AN INVESTIGATION
ON
WEB ENABELING OF LIBRARY MATERIAL IN OPEN LEARNING SYSTEM

THESIS

Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY (Ph.D.)

IN

LIBRARY & INFORMATION SCIENCE

Under the Superevision of

Pro. M.T.M. Khan
(Co-Guide)
Head, Department of Library
and Information Science
Bundelkhand University
JHANSI (U.P.)

Dr. T.A.V. Murthy
(Principal Guide)
Director, INFLIBNET
AHMEDABAG (Guj.)

By

AJAY PRATAP SINGH

DEPARTMENT OF LIBRARY & INFORMATION SCIENCE
BUNDELKHAND UNIVERSITY
JHANSI

2004

CONTENTS

Prof. T M Khan
Head, Deptt. Of Library and Inf.Science
Bundelkhand University
Jhansi (UP)

Dr. T A V Murthy
Director
INFLIBNET Centre
Ahmedabad 380 009

CERTIFICATE

This is to certify that the work embodied in the thesis entitled "An Investigation on Web Enabling of Library Material in Open Learning System" is submitted by Shri Ajay Pratap Singh for the award of the degree of Doctor of Philosophy (Ph.D) in Library and Information Science. It is a record of the bonafide research work carried out by him under our supervision and guidance. This work has not been submitted elsewhere for a degree/diploma in any form.

It is further certified that he has worked with su for the period required under the Ph.D. degree ordinance-7 of Bundelkhand University, Jhansi.

(Prof. M T M Khan)
Co.Guide

(Dr. T A V Murthy)
Principal Guide

Dr. T. A. V. MURTHY
Director
INFLIBNET CENTRE
UGC, Ahmedabad-9.

DECLARATION

I do hereby declare that the thesis entitled **"An Investigation on Web Enabeling of Library Material in Open Learning System"** submitted to Bundelkhand University, Jhansi, has not previously formed the basis for the award of any degree, diploma or other similar title or recognition. This work embodies the result of my original research and reflects advancement in this area.

Ajay P. Singh

Date:22/01/2004

(Ajay Pratap Sihgh)

ACKNOWLEDGEMENT

As the time draws near to complete the research work, I can not help but recollect about the people, who have enlightened me and given me memories to be cherished life long, during my course of investigation.

With immense pleasure I wish to express my profound sense of reverence and gratitude to Dr T A V Murthy, Director, INFLIBNET, Ahmadabad, Principal Guide, for his constructive criticisms, through out provoking comments and excellent counsel throughout the investigation and preparation of this manuscript.

My heartiest thanks are due for Prof M. T. M. Khan, Head, Inst of Library and Information Science, Bundelkhand University, Jhansi, Co-Guide of my research work for his inflicting advice, affectionate guidance, instilling the importance of learning, believing in myself and urging me to strive for higher goals.

My positive poignancy tries to reach Prof Ramesh Chandra, Hon'ble Vice-Chancellor of Bundelkhand University, Jhansi for providing all the necessary permission and help during the course of investigation.

Thanks are also due to Prof. O P Kandari, Pro-Vice Chancellor, Prof P N Srivastava and Prof. Dheer Singh, Director (SFS) for their affection and guidance during this investigation.

I gratefully acknowledge Mr V K Sinha, Registrar, Mr Kali Charan, Finance Officer, Mr D N Bajpayee, Dy Registrar, Mr Mani Ram Verma, Asst Registrar and Dr Kamlesh Sharma for the necessary help and assistance.

My simcere thanks are due to Prof D.P.Singh, Vice Chancellor, U.P. Rajarshi Tandon Open University, Allahabad, Prof P.B.Mangala, University of Delhi, Prof J.L.sardana, University of Delhi, Dr R.K.Singh, Librarian, Kota Open University, Dr Madhukar Shivale, Librarian, Yashwantrao Chavan Maharashtra Open University, Nasik,

Dr (Mrs) Neela Jagannathan, Librarian, Indira Gandhi National Open University, New Delhi for their moral support and help from time to time.

I gratefully acknowledge Mrs Prof M T M Khan and all family members for their constant moral support, help and homely atmosphere provided by them.

I seize the opportunity to express my heartful greatfullness towards Mrs Prof Dheer Singh, Dr Priyanka Tomar, Ms Monica Tomar, Ms Silky, Ms Neha and Mr Amit for their constant encouragement during the investigation.

My simcere thanks are due to Dr Shiv Ram Verma, Dr. V. P. Khare, Mrs J Sridevi, Dr Joginder Singh, Mr M P Singh, Ms Shalini Vyas, Ms Atia, Mr Om Prakash and all the students of Department of Library and Information Science for their moral support and help from time to time.

Words are lacking to express sincere appreciation for Prof A K Saxena, Dr Pankaj Atri, Dr Sunil Kabia, Dr Prateek Agarwal, Mr Sanjay Niboria, Dr D N Yadav, Mr Ashish Chandra, Mr Alok Verma, Mr Khusendra Borkar, Mr Shashi Kant Verma and all the faculty members of Computer Science department.

The help, support and affection shown by all the Head of the departments, faculty members of different departments and University staff members can not be acknowledged in words.

I am unable to find words for expressing my feelings for the love, affection, moral support and time to time help provided by my Mummy, Daddy, Brother Dr Anand Pratap Singh, Bhabhi ji, Mr Santosh Kumar Singh, Mr & Mrs Saurabh Srivastava, Dr D K Bhatt and Dear sisters Ritu and Kiran, who have always been the guiding lamp in my life.

All the help and facilities received from the University during the course of the study program is duly acknowledged.

Ajay P. Singh

(AJAY PRATAP SINGH)

PREFACE

Chapter 1 deals with the introductory part of the research. It discusses the actual problem which is being faced by librarians of open universities in India along with the scope of the research, frames the objectives carried out during the planning of the research, hypothesis designed to give the work a specific direction and presents the methodology which have been followed during the course of research for data collection, literature survey and preparation of thesis.

Chapter 2 deals with the area of distance education. It presents a descriptive and analytical study of several definitions given by different scholars; various philosophies of the distance education, and their difference with the traditional system of education alongwith it's prone and cons. At the same time it examines the role of libraries while delivering education to open learners/researchers through distance learning in open universities.

Chapter 3 deals with the development of various mark up languages, world wide web consortium's standards, web designing tools and various technologies supporting web page designing for a library's portal. While discussing web technologies it is kept in mind that the requirements of open learners should be the core criteria for taking any decision.

Chapter 4 deals with the digitization process. It examines various important points to keep in mind while digitizing a document. Some of these are pattern of digitization of text, pictures, digitization formats, copyright issues, digitization tools and techniques. It also examines an ideal pattern of materializing a digitization project. In the last part various benefits of digitized document have been discussed over document in text form.

Chapter 5 deals with the comparative analysis of various telecommunications technologies for data communication over world wide web. It discusses some of the common web server technologies along with their standards, patterns and features, etc.

Chapter 6 deals with the analysis of data. The data is collected from the students studying in the various courses of selected open universities in India. Data is collected through a well-designed questionnaire, interview based on interview schedule and personal discussions during the visit of the researcher to open universities and their study centers. Data is analyzed in the form of tables and graphs and later on narrated for generalizing the result.

Chapter 7 deals with the findings of the research. It is presented in the form of various points. It is based on the analysis of the data collected during the course of research. It also presents the limitations of the work done and the further area of research may be taken place in the future.

First Annexure is a bibliography of those books or documentary sources that have been consulted as the study or reference material during the course of research.

Second Annexure is a list of those web sites that have been consulted during the course of research.

Third Annexure presents the Questionnaire that has been distributed among distance learners during the course of data collection. This questionnaire basically deals with their expectations from the libraries of their respective open university.

CHAPTER 1 : INTRODUCTION

- INTRODUCTION
- PROBLEM
- SCOPE OF THE STUDY
- OBJECTIVE OF THE STUDY
- LITERATURE SURVEY
- HYPOTHESIS
- METHODOLOGY

1.1 INTRODUCTION:

With the advent of more and more powerful information technology tools, the role of information specialists become complicated, because of his diversified information requirements and new ways of information dissemination. As libraries are an integral part of the Information Super highway, we must develop a library that fit the "World of Tomorrow". By the invention of different networks critical milestones have been reached, but we information professionals are at an important juncture.

The library world is changing rapidly. Few years back we had libraries with books and other material in paper form. Later on, we switched over to computerized library and then automated libraries. At the same time we use to procure study material in digital format, so in present scenario we have different scenes in our single library, i.e. library with study material in traditional paper form, library with study material in digital form or paperless library and we use to manage it side by side separately.

1.2 PROBLEM:

Open learning system provides education and training to all those persons, who are unable to learn full time while being present physically. Majority of persons has their own limitation resulting them to join open learning system. As an open learner rarely comes to the campus, it becomes essential to provide study material at his end. Now a days open learning centers are providing such facilities efficiently. But this study material cannot fulfil~~l~~ their curiosities and information needs. So ~~that~~ they need some other related or reference material. In practice it is not feasible to provide library material to these open learners at their end. This problem becomes more serious, when student becomes a researcher. For providing some solution of the above-mentioned problem, the present study investigates the ways of the web enabling of such library material for a distant learner.

1.3 SCOPE OF THE STUDY:

The present study is confined to investigate various ways of:

- Designing a good library web page and
- Digitizing library documents

For this purpose it is required to survey the basic requirements of open learners of some of the prominent Indian Open Universities. It is also required to find out various techniques for Web Designing and Digitization and to find out the most appropriates out of them.

While analyzing the information requirements of Open Learners, it is decided to visit Indira Gandhi National Open University, New Delhi; Yashavantrao Chavan Maharashtra Open University, Nashik; Kota Open University, Kota and Rajarsi Purshottam Das Tandon Open University, Allahabad.

The main thrust of this study is to make library documents available through web in open learning system.

1.4 OBJECTIVE OF THE STUDY:

The objectives of the present research work may be enumerated as under:

- To review the existing status of open learning system.
- To identify the information needs of beneficiaries of open learning system.
- To investigate on the conversion of Library material in web enabled form.

- To identify the basic S/W and H/W requirements.

- To investigate on the library accessable through Web.

1.5 LITERATURE SURVEY:

The emergence of new information heandling technologies have significantly influenced the basic nature of conventional paper based libraries and have created a need for a new type of library systems as polymedia, electronic, digital, and virtual libraries. (Barker, 1996). The popularity of e-Books has grown since their inception in the early 1980s due to their usefullness in distributing large volumes of interactive multimedia information. (Barker, 1996). Barker (1996) has reported the basic nature of eBooks, the philosophy underlying their use, the basic texonomy and description of various techniques involved in their design and fabrication. A comprehansive media strategy allows information to be moved from one medium to another as the needs of its users change. (Barker, 1998). Landoni et al (1993) reported two innovative forms of eBooks as hyperbooks and visual books that are based on the book metaphor and the environments, in which such eBooks are produced. Roberts (1999) describes how an academic library provides dynamic access to ever changing serials holdings. Roberts (1999) again describes a web based database containing ready reference sources, unlike many library sites in which reference sources are hard coded as links on a web page. Ervin (2000) describes how the Jackson library at the University of North Carolina converted a directory of online news papers from static HTML files to a Microsoft Access database and then delivers the requested information using Active Server Page technology. A good web based tutorial on Common Gateway Interface is designed by Selena Sol (1998).

The Z39.50 Information retrieval Protocol has provided the facilities for automating information systems and bibliographic databases. Traditional libraries face space and financial restrictions, since the amount of holdings rapidly expand with the cost of individual publications while library budgets are continiously decreased. (Luther, 1999).

The phrase "Electronic Document Delivary System" (EDD Systems) self evidently implies the supply and reproduction electronically of the kind of information usually provided in the form of print on paper. The three generations of EDD systems can currently be distinguished: systems based on online ordering, non-integrated supply-driven image-based systems, and integrated stand-alone image-based system. (Roes and Dijkstra, 1994). Critical EDD system technologies are not yet adequately developed and most publishers still publish printed materials more than any other material. The basic reasons of delay are examined by Berghel. (Berghel, 1999). Some of the EDD systems are NAILDD project (Barrett and Jackson, 1993) and ARIADNE system (Roes and Dijkstra, 1994).

Metadata, a fundamental role of digital content, has now become an important part of the globle information construction in planning, processing, restoring and managing. (Vellucci, 2000). Vellucci has also listed a number of metadata sets. (Vellucci, 2000). There are more than 20 different types of international standard metadata existing among the domains for different requirments. (El-Sherbini, 2001). It is ideal to establish a higher level of super metadata for all metadata interoperability. It facilitates the success in integration, and each metadata keeps its on character. (Chilvers and Feather, 1998).

ALA affirms the right of all persons to access electronic information in its interpretation of the library bill of Rights by stating that "Electronic information services and networks provided directly or indirectly by the library should be equally, readily, and equitably accessable to all library users. (American Library Association, 2000).

The World Wide Web has rapidly become the most popular internet resource, combining hypertext and multimeia o provide a huge network of educational, governmental, and commercial sources. (Burgstahler et al, 1997). WWW is one of the tools that uses the hypertext and allows computers to link information in new ways different from a sequential reading approach, to make it easy to retrieve and add information from different computer sources through the use of computer links. (Berners

Lee et al, 1992). Web will bring forth a better democracy within the USA by returning the power to the people. (Meeks, 1997). The number of users from medium income group joining the web is higher then the number of users from higher income group. (Pitkow, 1996). One of the lacuna of internet is that of inadequate search facilities with the lach of a high level quary language for locating, filtering and presenting WWW information. (Foo and Lim, 1997). It is difficult to locate a desired web site by majority of users. (Pitkow, 1996). In case of web site maintainance and assurance of information accuracy is difficult. (Foo and Lim, 1997). Many systems allows software developers to attach programs which are executed upon access to a web page. It is called webware "simply visiting a web page may cause you to unknowingly down load and run a program written by someone you don't know and don't trust. (Felton, 1997).

The World Wide Web currently has a huge amount of data with perticularly no classification information and this makes it extremely difficult to handle data/information effctively. (Marchiori, 1998). The task of knowledge management can be accomplished by adding to web objects a metadata classifiction which will assist search engines and web based digital libraries to proprly classify and structure the information on the WWW. Bartlett (1999) points out that accessible web sites allow web search engines to more effectively index web pages. (Marchiori, 1998). Bartlett (1999) also states that the web is not exclusively visual medium, but rather an information medium; one way to convey that information is visual. He goes on to comment that Hypertext Markup Language (HTML) is designed to display content independent of a specific means of representation and that web page creators who only design visual pages are missing out on the power of HTML. Accessible web page also allow the optimum use of screen reading software and other adaptive computer equipment. (Coombs, 2000; Cunningham and Coombs, 1997). People with visual disabilities may experience low vision, functional vision, color blindness or blindness and have problem seeing computer screens and using keyboards. (Cunningham and Coombs, 1997). Persons with physical or motor disabilities may often be unable to use standard computer input and output devices. (Cunningham and Coombs, 1997). Web designers may reduce barriers to access by implementing a simple design

that is easily viewed and incorporating clear on screen and keyboard navigation. (Coombs, 2000). Those with learning disabilities may have visual perception problems and/or aural processing difficulties. (Cunningham and Coombs, 1997).

1.6 HYPOTHESIS:

To achieve the specified objectives of the present study, following hypothesis have been formulated:

1. It is possible to convert library documents in digital format.

2. It is possible to make available the digitally stored documents through web.

3. Web enabled library material can efficiently and effectively fulfill the library needs of a open learner.

1.7 METHODOLOGY:

This investigation has been carried out using personal computer equipped with various tools and techniques for digitization and web page designing. Major steps involved in the methodology are given as under:

Step 1	Literature Search
Step 2	To conduct a survey for identifying the information requirements of open learners.
Step 3	Finding out appropriate technique for designing library web page.
Step 4	Finding out appropriate technique for digitization of documents.
Step 5	Investigation on the enabling of digitized documents on web.

During the course of selection of the area of study / research, an extensive search

of literature had been carried out. Various bibliographic tools had been used.

At the time of starting research in the predefined area, various related areas of study have been sorted out. Related literarute on these sub areas have been searched to make the vision clear. Some of these major areas are:

Digitisation,

Web Designing Technology,

Web Server Technology,

Language for Web Page Designing,

Distance Eucation,

Open Learning System in India, etc.

In the second stage it is tried to identify the actual learning conditions and information requirments of Open Learners. It was done by designing and distributing questionnaire for open learners. During this stage, it is required to go through direct interaction with the related persons. For this a survey was conducted to interview open learners and distance educators.

In the next stage, various tools and techniques have been examined to search best out of them for the purpose of Digitisation and Designing of Library Portal. For this purpose investigations have been carried out in the computer laboratories of Bundelkhand University, Jhansi and Information and Library Network (INFLIBNET), Ahmedabad.

In the fifth and final stage all of the investigation has been combined and grouped together to conceptualise an ideal library having digitised material and accessible through World Wide Web.

CHAPTER 2 : LIBRARIES IN OPEN LEARNING SYSTEM

- **DISTANCE EDUCATION**
- **LIBRARIES IN OPEN LEARNING SYSTEM**
- **IMPROVEMENT OF THE LIBRARY SERVICES TO MEET THE NEEDS OF DISTANCE EDUCATION**
- **INTEGRATED INFORMATION SYSTEM FOR OPEN LEARNING**
- **LIBRARY NETWORK OF OPEN UNIVERSITY IN INDIA**
- **CONCLUSION**

2.1 DISTANCE EDUCATION:

The educational system of the past was of highly elitist in nature. In such system education was confined to a few dominant groups in the society. The Greeks or the Romans have never allowed any freedom to the learners, who were considered as passive agents ready to receive whatever the teacher choose to give them. In Europe, the Church always controlled education. The Church supported class structure in the society and it taught those ideas that were conducive to the teachings of Christianity. In India, Brahmins dominated the society and they did not allow the lower castes to be educated. In India, the pre-independence English system of education catered the needs of the English Rulers and theirs supporters. It was aimed at supplementing the erstwhile Oriental Education. After independence we thought to have snatched the educational supremacy in India from the British by taking up a massive program of education.

While analyzing the elitist nature, main disadvantages of education are:

- It is restrictive in nature. Only a group is allowed to be educated. So it does not provide access to education to a large section of the society.
- It does not allow learner's autonomy.
- Education does not help in social transformations.
- It does not allow flexibility in the educational system.

To remove these ills of elitist concept of education democratization of education has been advocated.

Main reasons that were demanding the need of higher education in India are:

- Population explosion,
- Economic development,
- Social transformation,
- Desire for a white – collar job.

The Socialistic principle of our constitution promises equal opportunities for all. So various education commissions of India advocated universalisation of education. There are some mismatches between the Socio-academic needs of our society and the conventional educational pattern. This mismatch was not striking easily in the past as societies evolved slowly and therefore absorbed educational products. But later on society started rejecting the products of institutionalized education. As they failed to solve real world needs. Our Socio-academic needs are as follows:

- Need of part time education with a flexible arrangement in order to meet the requirements of young persons who can earn while learning,
- Need of specialized courses for these who are in service,
- Need of intellectual stimulation for adults,
- Need for certificates/diplomas beyond the scope of the conventional university system.

Our conventional university system is not adequate to cope up with these needs because:

- They do not allow earning while learning. Correspondence courses are the exceptions but they do not provide teaching aids,
- The highest paid teachers are catering to the needs of only a few students,
- The age old face to face teaching situation is becoming ineffective,
- There is rigidity regarding duration, timing, attendance etc.

Besides the conventional education system if we analyze the attitude of present day learners, we found a markedly different attitude from their predecessors in the following aspects:

- They like to follow novel ideas,
- They do not want to get into a pre-conceived and pre-planned educational system, they are masters of their own mind,
- They want to be independent of their teachers,
- They have a belief in their ability to change the society for better,
- They want flexibility in the educational system.

Analyzing these aspirations of new learners carefully, we realize that the conventional system of education is not able to fulfill all of them. Let us examine how far distance education is able to satisfy the new learners:

- Distance education emphasizes learning rather than teaching. So the onus of the system lies on the learners rather than teachers. The teacher merely acts as a facilitator of learning. The distant learner learns at his own time and own pace,
- Distance education provides many non-conventional courses, which were earlier missing from the curriculum of a university,
- Human being is not empty vessels that can be filled-in with knowledge, so the learning process is bound to determine by what the learner can take and how. Distance education provides enough flexibility for this.
- Distance education allows the learners to be autonomous. He feels his own progress with passing of time, if he continues in distance education; the learner chooses his own course of study that makes him responsible for the relevance of what he is studying.

Thus, it is sum up that after examining the ills of the conventional system of education and present day learner's attitude, it is fined that there is a mismatch between the socio-academic needs and the educational assumptions. Besides this the aspirations of the new learner are not fulfilled by the present system of education. That is why institutionalized education is gradually losing ground. Hence the solution to the quest lies in distance education.

To overcome this mismatch, Distance education can be used effectively. It can provide not only conventional courses in a capsular from but also allow individual variation in it. Moreover, it can provide a variety of post experiences and in service courses that are more geared to the requirements of the society and adults. In other words, distance education can provide conventional education and continuing life-long education side by side. In U.S.A., Canada and other western countries it is a successful tool for life

long learning. It has overcome the mismatches that exist between the Socio-academic needs and the conventional educational assumptions.

Here are some distinct advantages of distance education over the present system of education:

1- This system is able to cater to the educational aspirations of innumerable aspirants, who are otherwise deprived of their education. All willing persons, regardless their age, sex, employment nature, place of residence or social status can join these courses to improve their sphere of knowledge.

2- Capital expenditure incurred for distance education is relatively economical. The amount required to cater to 1000 students in the face-to-face teaching situation can prove sufficient to teach many times this number through distance learning. This factor has been of prime importance to developing countries like India.

3- The flexibility of distance education is another great advantage. It allows greater freedom to the learner. A variety of courses can be offered through it. A particular course, if proved unpopular, can be withdrawn without much waste.

4- It allows students to earn while learning. Distance learners can employ themselves and study in their leisure at their own pace.

5- Student unrest is conspicuously absent in the distance education institutions. Student indiscipline, campus violence, which is common features of educational scene, is never a problem for the administrators of distance education.

6- The defence personnel of a country can improve their training and education through distance learning. These men in uniform often do not get a chance to enhance their qualification or training while in the defence job. But distance education can provide such retraining.

Distance education has become very successful tool in the western countries in fulfilling the educational needs of millions of learners. But it is not possible to borrow a model of distance education from any one of these countries. It is because of the socio-academic, economic and cultural differences among developed and developing countries. Here it is needed to study various existing western systems of distance education closely, mark their advantages and disadvantages; and then evolve a new system of its own.

2.11 DISTANCE EDUCATION: PHILOSOPHIES AND THEORIES:

The term distance education has come to be associated with non-conventional teaching or learning programmes, where teacher learners build their relation mainly through printed words instead of oral instructions. In case of conventional system of education, the onus of teaching lies on the teacher. The methods and the materials used for teaching are geared to the end. Only recently it has been shown that learning is possible through written words also. The teacher through his words of mouth communicates his own ideas, interpretation etc. But such communication can be possible even through means other than words of mouth. When the need arose, students living far from their teacher also learn effectively. In 1840, short hand could be taught through post and in course of time all types of courses were taught through postal correspondence. Even engineering courses were being taught through correspondence in the erstwhile USSR. Hence the name 'Correspondence Course' or 'Correspondence Study' stuck to all non-formal channel of education. Only recently the term distance education has been coined to embrace all programmes like Home Study, Postal Tuition, Correspondence Courses, Non-formal Education. The world body for correspondence education has also changed its name to the International Council for Distance Education from the International Council for Correspondence Education.

'Distance education' is an improvement over the term Correspondence Education as it is supposed to be an improvement over the aims, methods and approaches of Correspondence Education. Whereas Correspondence Education depends mostly on printed materials for teaching, Distance Education employs multimedia approach including postal system.

The basic philosophy of distance education can be cited as follows:

- No one is too old to learn,
- Education is a life long process,
- No one is master enough to shun new ideas, methods and concepts,
- Even without being admitted to a school/college, one is not barred from education.

While defining Distance Education, the major theoretical formulation of distance education so far published is Otto Peter's Die didaktische Struktur des Fernunterrichts. Otto Peter[i] was associated with the Deutsches Institut for Fernstudien an der Universtat Tubingen. Later he becomes foundation president of the Fern-Universitat-Gesamthochschule in Hagen. He says about distance education as "Distance education is a form of indirect instruction. It is imparted by technical media such as correspondence, printed materials, teaching and learning aids, audio visual aids, radio, television and computers". He concluded that "the didactical structure of distance education can best be understood from industrial principles especially those of productivity, division of labor and mass production". Peter attempted to define the relationship between the teachers and taught in a distance education system. He characterizes this relationship as being controlled by technological rules (and not social norms as in face to face teaching), maintained by emotion free language (and not interaction speech), based on a limited possibility of analyzing student's needs and giving them directions (not on expectations built on personal contact) and achieving its goals by efficiency (and not through personal interaction).

Besides this some other theories of distance education given by eminent distance educators are summarized as follows:

Borje Holmberg[ii], a Sweden born and federal republic of Germany national define it with two elements of special consideration as

- The separation of teacher and learner,
- The planning of an educational organization.

According to him the separation of teacher and learner is fundamental to all forms of distance education whether they be print based or audio radio based, video television based, computer based or satellite based. This separation differentiates distance education from all forms of conservational face-to-face, direct teaching and learning.

French government passed a law regulating the conduct of distance education in its territories on 12th July 1971. This gives special emphasis to following two points as

- The separation of teacher and learner,
- The possibility of occasional seminars of meetings between students and teachers.

Michel Moore[iii](1977), a senior counselor in the southern region of the Open University of the United Kingdom has worked extensively in the United States of America. Main elements of his definition are:

- The separation of teaching behaviors and learning behaviors,
- The use of technical media,
- The possibility of two-way communication.

In his opinion, in distance teaching system, the teaching behaviors are performed apart from learning behaviors. Yet communication between the teacher and the learner

are facilitated by various media like print, electronic etc. He is of the opinion that the interaction between the learner and the teacher determines the efficacy of the system.

These three definitions so far considered can probably be accommodated within any basic theory of education. *

While discussing it in the context of India Education Commission[iv], 1966 says "Education is most important single factor in achieving rapid economic development and technological progress and in setting social order founded on the values of freedom, social justice and equal opportunities." Hence, it recommended correspondence courses, popularly known as distance education, as an alternative to the conventional system of education.

2.12 DISTANCE EDUCATION AND NEW TECHNOLOGIES:

Distance education cannot insulate from new technological imperatives. That is why a growing interest and an emotional fascination with the use of modern communication media is found in distance education.

But only a few institutions use the media in a significant and substantial way. Some Australian Universities use it as an alternative to print media. The open universities in United Kingdom use radio, audiocassettes, television, broadcasts and videocassettes as components of study material. Some institutions in Europe, Canada and United States use videocassettes along with print. Technologies should be used on considerations like weather the use is administratively convenient, financially viable, technically possible, pedagogically significant and accessible to the student users or not.

The National Policy on Education clearly recognizes the role of technology in open learning as "The open university system has been initiated in order to augment opportunities for higher education and as an instrument of democratizing education"(NPE, 1986). Referring to educational technology, the policy document observes, "Modern communication technologies have the potential to bypass several

stages and sequences in the process of development of time and distance which at once becomes manageable. In order to avoid structural dualism, modern educational technology must simultaneously reach out to the most deprived in the most distant areas and comparatively well off in area of affluence and ready availability.

2.13 CHARACTERISTICS OF DISTANCE EDUCATION:

Some of the major characteristics of distance education are as follows:

1. Teacher and the taught remain separated in this system: There is quasi-permanent separation of teacher and learner throughout the length of the learning process. Very often geographical barriers separate the teacher and learner. They have a chance of meeting each other during personal contact program, but such meetings are few and far between.

2. Learning is very individualized in this system: There is a quasi-permanent separation of learner from his peer group throughout the length of the learning process. Distant learners do not even know each other. They live scattered at various places, learning at their own places of living and at their own pace.

3. Oral communication is replaced by multimedia technology in distance education. Many media are used like printing, telephone, audio-video tapes, broad casting and computers. Personal contact classes are also held to add a personal touch to the system. These additional media are used to reinforce learning.

4. In this system, the onus of teaching lies on the institution and the onus of learning lies on the learner. Distance education is an institutionalized system of education, which distinguishes it from private study. According to Erdos[v], in correspondence education, "...teaching responsibility lies on the part of the institution. It is a method of teaching in which the teacher bears the responsibility of imparting knowledge and skill, to a student... who studies in a place and at a time determined by his individual circumstances."

5. Two-way communication between both student and teacher is possible in distance education. In an otherwise mechanized system of education, this two-way communication brings in a touch of fresh air. Students assignments and there after teacher's comments, suggestions for further improvement make up for the loss of personal touch. Therefore many western distant educators give importance to this aspect of distance education for an effective distance teaching.

6. Another important characteristic of distance education is the industrialized way in which distance education is organized. It is true that distance education has arisen from the needs of an industrialized world. The working of an institute of distance education also resembles an industry because of mass production of study materials, divisions of labor in the institute and the layout of the institute buildings.

7. Distance education is very democratic in the sense that it is open to public inspection and criticism. A face-to-face teaching situation is basically private in many ways. Oral communication is restricted to the classroom or a group of students. It is not open to everybody for review; on the other hand, the study materials provided in the system of distance education is seen criticized, revised and reviewed from time to time.

2.2 LIBRARIES IN OPEN LEARNING SYSTEM:

2.21 ROLE OF EXISTING LIBRARIES IN DISTANCE EDUCATION:

Libraries are always concerned about the needs and demands of their users. In India, more than thirty Universities are conducting Correspondence Courses and with the

establishment of national and state Open Universities, the number of user communities consisting of mature distance learners is increasing enormuously. Again, their requirements are diverse. Therefore, libraries of all types viz., academic, special and public should focus their attention on meeting the library and information needs of independent adult students.

As far as the academic libraries are concerned, most of the school and college libraries in India exist only in name. A few well organised libraries having adequate collection do not possess enough infrastructural facilities to cater to the needs of their ow n students. So the question of providing library services to a mass of distance learners does not arise.

2.22 ROLE OF DIFFERENT TYPES OF LIBRARIES:

2.221 Public Libraries:

Public libranes are often called the university of the public. They have an important role to play to make distance education a success. It is the public library where any citizen can get a membership to use the library. There are restrictions in other types of libraries. The public libraries can act as the regional centres easily if slight modification is made with the assistance of distance education system. The local libraries like the municipal, panchayat, and village libraries can very well act as the local centres of the distance education programmes.

A person can undergo a course under distance education programme with 'learning' as the emphasis without a tutor. This fits in well with the public library since it also serves all without any restriction. The subject of distance lerning has been gathering momentum in recent years and libraries have to be involved in the programme in a developing role. Among the libraries, the public library, a vital link to learning, has a greater scope to motivate adult to learn and to attain their learning activities by having

independent study. As such the public libraries have to become fully involved in Distance learning since the learners are not able to use acadenuc or special libraries.

The public libraries, therefore have tB bring together the potentialleamers of their region and their relevant materials and should link for maximum usefulness the learners as part of its regular service and prepare itself for these new roles as a commitment. This would be in line with the national educational policy of the government as part of its total pattern of service to the community. Thus the public library is the one institution accessible to all and able to cater for the interests of all.

In Britain, the main purpose of United Kingdom Public Library system is to promote self-education. The British Open University does not have libraries at their Regional and Study Centres. The students of U.K.Open University depend mainly upon public library services for their independent studies. Lord Walter Perry, the first Vice-Chancellor of the British Open University observes in his book 'Open University':

"As far as the students were concerned, scattered as they were throughout the whole of the country, it would not be feasible to offer a library service. They would have to rely on the public libraries and on inter library loan services to acquire the reading material that they would need".

In Thailand, students of the Sukhothai Thammathirat Open University and the Ramkhamhaeng University usually avail of library services at the nearby public libraries. The public library systems in these countries have attained great popularity for providing services to the adult learners.

2.222 Academic Libraries:

2.2221 University Libraries:

University libraries are at present allowing the public to a limited extent to use their resources. By slightly relaxing the, condition, the university libraries can be turned,

into highly resourceful centres: We know that there is the need for local and regional centres and local guides to make the distance education a success. University and public libraries are the right places if put into right use. The resources of these libraries include documents and professional library staff. The library staff can easily guide the students of distance education acting as their local guides. It is appropriate and economical to appoint professional library staff with qualifications in different disciplines other than library and information science. These library staff can be trained easily as guides. This programme can be implemented in a University library. There is na doubt that the knowledge of library and information science plus other discipline will prove to be more fruitful in guiding the students.

2.223 Special Libraries:

We always keep special libraries out of our ,discussion on education. Though the Medical College, Engineering College and IIT libraries can be designed as special libraries, our stress here is the real special library which is part of an industrial or research establishment. These special libraries are also playing an important role in formal and distance education. Distance education is an opportunity to all for pursuing studies at any time. A working person, say, a scientist or engineer has a chance to continue his study while working. They are using mainly special library resources for their study purposes. Mostly these studies are useful to the organization concerned and the special libraries should pay attention to their needs. The present trend in distance education shows that science and technical courses can also be conducted through distance learning.

2.23 REQUIRMENTS OF OPEN LEARNERS:

The requirements of the distance learners can be grouped in three categories.

(i) need for materials & facilities;
(ii) need for information services; and
(iii) need for user services.

2.231 Need for materials and facilities:

Distance learners should have facilities for consulting library materials on the premises, borrowing them and getting them on inter library loan from the other libraries through public libraries. The selection and acquisition of self instructional materials and open learning packages developed by various distance teaching universities should be given priority to develop the user-oriented collection. The provision to procure audio-visual and non-book materials from various organisations should be made to facilitate self-learning. Public libraries should be well-equipped with audio-visual equipment and hardware. They should have lecture room/ discussion room for the independent learners to meet and to view/hear audio-video cassettes.

2.232 Need for Information Services:

The staff of the public libraries should be trained to collect and retrieve information for the distance learners as and when required. Even if the materials are not available at some small public libraries, they should be able to provide up-to-date infonnation on the following:

(i) Bibliographical information of reference source and tools, books, journals and other print materials available in the library;

(ii) Bibliographical details of materials available in other libraries which can be borrowed on inter library loan;

(iii) Infonnation regarding various self-instructional or open learning packages of several open universities in India and abroad;

(iv) Infonnation of audio-visual and multi-media materials and their availability;

(v) Information regarding eucational programmes of radio broadcasts, television telecasts and related materials about them;

(vi) Information regarding the organisations imparting education and training opportunities through distance teaching and their various courses; and

(vii) Information regarding Regional & Study Centres of various organisations, timings of contact and counselling programmes, summer schools, special lectures, laboratory workshops.

2.233 Need for User Services:

Public libraries have a special responsibility to provide user services to adult distance learners who need professional guidance and support in-

(a) Using the library collection;

(b) Selecting the reading materials;

(c) Planning their learning; and

(d) Utilising the study skills for self learning.

2.24 SPECIFIED SERVICES PROVIDED BY LIBRARIES:

2.241 Building up libray collections specifically for the purpose of distance education:

As has already been pointed out; in distance education, there is a shift from teacher-centred to learner-centred educational system. Students have to rely more and more upon themselves. Most adult students are busy people, often with heavy occupational and domestic responsibilities. They naturally expect to have a library service near at hand if they are to take full advantage of it. Academic libraries can ensure that the needs of students are properly provided for and the most effective way of

meeting these needs is from collections of books and other materials specifically built up for this purpose. Christopher Barnett goes a step further by saying that even public libraries can be of enormous help to such students. Libraries can also launch 'Learn with Your Library' programmes for the benefit of students.

2.242 Postal Library Service:

Further, the students in this type of education (who may appropriately be described as off campus students) may place special demands on borrowing systems and these demands should be full filled with the provision of outreach. After all, as Haymond Fisher has pointed out, "if you are teaching students at a distance, you must send the resources to them, precisely because by definition they are separated from the main provision of the library". With the aid of postal library services, libraries can make available books and other materials required by such students.

2.243 Learner's Advisory Service:

Libraries cari also provide learners' advisory services to these students. By acting as can adviser and facilitator, the librarian can provide support to such students by means of information.

2.244 User Education Programmes:

Libraries can also take up user education programmes. These programmes will help users. understand library systems and layout. Retrieval of library documents in such cases will be integrated with learning programmes.

Now it is clear that libraries can play a much important role in distance education. In the words of C.S.Hannabuss, "In a so-called integrated teaching and learning system, the library can no longer be merely a collection of printed materials. It has to be an instructional resource centre, handling a wide range of print and non-print resources arid guiding independent inquiry.

2.245 General Career Guidance Information Centre:

Organisation of a General Career Guidance Information Centre (GCGIC) which provides information on general career guidance will be fruitful. Such information can be delivered at the counter as well as through correspondence. In addition to the basic information on sciep.tific career, the comprehensive career centre also includes guidance counsellors; opportunities to take tests of skills, ability and creativity. Emphasis should be laid on machine oriented careers, because maximum students of any open edut'lltion system would be common masses, workmen etc. who could not avail of the various opportunities of conventional education because of various constraints.

To be fully functional (GCGIC) should have the support of the entire community. It is not sufficient to have a guidance/counselling staff and enthusiastic support of the centre among library personnel. The entire community, faculty, administrators, students, and the public must be aware of the centre, its purposes and needs. The best publicity for initiating the use and for sustaining the value of CCGIC depends on the appropriate materials, being available, easy access and competent service.

2.246 Continuing Mass Education:

DELS (Distance Education Library System) can help the common masses to educate themselves continuously. DELS can also help millions of people to move forward in vocational and proffessional skills. DELS can also serve as an agency for eradicating illiteracy among the masses. A good collection of books and other reading materials and a well trained staff can exert a powerful influence against illiteracy as well as providing opportunities for the common masses to continue their education.

2.247 Extension Activities:

Extension activity may be defined as special library activities which are undertaken with the object of reaching groups of people who might otherwise be unaware of the library, such as lecture groups, reading circles, discussion societies etc. It may also mean provision of lectures, film shows, etc. arranging talks and book displays within the library or outside the library. However, in case of DELS, the extension activities can be

organised for stimulating the reading interest and publicising the functions and services of the system as a whole. Modem techniques of communication such as Radio, TV, Video Cassettes, Audio Cassettes, Film Shows etc. are also advocated under the extension library service. The lectures and group discussions may be organised in DELS directly or in association with some institutions and specialist speakers should be invited to speak on various developments in selected topics. Displays and exhibition of books and other reading material is now considered t'obe an essential part of librarianship.

2.248 Mobile Units:

A mobile Unit library is a stock of books kept in a vehicle with limited staff to provide library service to scattered communities and providing in some cases, house to house service in remote areas such as villages and hamlets. In a country where majority of the population live in villages, this type of extension.activity is most important as well as useful. Bringing books to people by book mobile is a most dramatic as well as colourful type of library service.

2.249 Libraries at Study Centres:

Organisation of important library services in various study centres scattered all over the distant areas and the organisation of full-fledged study centres is a difficult task. But the particular distance education system can select some college libraries or district libraries for rendering the various library Services to various types of students residing in the territorial jurisdiction of a particular college or district library. The Library attached to the distance education system should finance these libraries, to the extent possible, so that these can be utilised as study centres, for its readers.

However, so far as possible distance education system should establish its own study centres in various localities and places including the rural areas. The establishment of one study centre for about 2,000 students would not at all be a costly affair. All these study centres should work under the control of the librarian or Director of Distance Education Library System (DELS). And these study Centres should open during evening

hours, because majority of the students of any DES would not be able to attend these study centres during normal working hours, because of being pre-occupied during the day time in the various occupations.

2.3 IMPROVEMENT OF THE LIBRARY SERVICES TO MEET THE NEEDS OF DISTANCE EDUCATION:

The following are certain suggestions for improving the library set up to suit'the requirements of distance education:

2.31 **Library legislation:** To strengthen the public library system.

2.32 **User education:** The students of distance education should be well trained in using the library. For this purpose, public and University libraries should plan and implement user education programmes.

2.33 **User Survey:** The public and university library should conduct user survey to get an idea about the requirements and aruilysefue feed back for suitable modifications.

2.34 **User assistance:** Libraries must activate reference service and user assistance programmes as the students of distance education may need help to locate information and complete their assignments.

2.35 **Application of modern technology:** Libraries should be equipped with facilities like satellite communication (Satellite TV), computer networks, videos, cassette recorders, rnicroforrns and others.

2.36 **Manpower development:** In order to act as local guides the libraries should appoint professionally qualified librarians from different disciplines. They are to be trained in the field of distance education.

2.37 **Library extension services:** Library extension services can be started or strengthened to incl ude seminars, talks, symposia, exhibitions, on topics relevant to distance educatiqn courses and mobile libraries.

2.38 **Finance:** Finance is always a problem in universities. Extra services mean more expenditure. The Open Universities and the UGC should take these into consideration and provide timely financial help to enrich the resources of the university and public libraries.

2.391 **Library Network:** As a future programme planning may be done to form a university and college library network. This can be on,the lines of online computer library centre (OCLC). This implies compilation of union catalogue, co-operative acquisition, etc. which would be very useful for sharing library resources.

2.392 **Library education:** The distance education curriculum for all courses should cover a topic on how to use libraries, this will help better utilization of library resources.

2.4 INTEGRATED INFORMATION SYSTEM FOR OPEN LEARNING:

We are in the era of an information society and its main thrust is on information management, the acquisition, processing and instantaneous, dissemination of information is the order of the day. I am sure you will agree that information science and information processing industry are a very important part of the basic production factors of the economy like land, labour, Capital and any investment in this sector pay off in present as well as in the future in terms of escalation of Research, Development of Technology employment generation, etc.

It is proper that the Government of India gives all support for these services in a

sizeable way and in particular give financial assistance, 100 percent for augmenting the public libraries text-book collections. In the information society, the concept, functions and structures of library will also undergo dramatic changes. Unlike today the libraries oi"tuture may in addition to stocking printed books to a certain extent will also stock software packages, information modules, etc. In the changed functions of public libraries micro processerswill find ever increasing applications. It is in fitness of things that we consider these applications the present day context and come out with suggestion which atleast some of our public libraries could adopt.

The problems involved in the implementation of an information system can be viewed from a number of different levels as being of relevance in the design of educational simulation or library management (lones 1984). National level, that is the educational setting within higher education involves the objectives expressed by the Government and bodies suchs as the U.G.C This consists of broad policy issues as also detailed policy statements. The institutional level involves the objectives of individual institutions, their organisation, structures for curriculum development and resource allocation, the role of library and other service departments, and course organisation procedures. The departmental or school level must take account of the objectives of individual departments and their organisation, structures for curriculum development and resource allocation, course organisation procedures, the educational climate of the department (including the hidden agenda) and relationships with external bodies (for example professional organisations and employers).

The development of open learning/ distance education areas and their associated information system requires a suitable climate at all these different levels within the educational world and also suitable resoruce allocatioon. In most distance educational institutions in India, otherwise known as Institutes of Correspondence courses, it is disappointing to find that they do not possess the capabiliites for use of the library information system nor do they have the attributes mentioned above. In some cases library informatiorl sub system has been developed in such a manner as to give the

impression that they reinvent the wheel oftraditional infrmatioon systems. The development of informatioon systems to support open learning/ distance education, therefore, requires an appropriate educational climate, funding and also a cross fertilisation of ideas from the worlds of librarianship and information science, computing, education and psychology. The practical implementation of a suitable system would require the selection of an appropriate set of open learning courses to provide the appropriate educational setting and resoruces. The development of the ideal library information system for open learning would require highly qualified staff in the field of librarianship. The establishtnent of an expert library information system requires much work and original research and perhaps may involve bigger pitfalls, which are harder and more expensive to climb out of. As a final word, I may add that it is possible to specify the ideal requirements of a library information system in an open learning/distance education situation as I was trying to do. Such a system would by and large fulfil information and educational objectives. The practical problems involved in the develpment and implementation of such a system appear to be related to the educational policy and setting at various levels and also to the provision of adequate resoruces.

2.5 LIBRARY NETWORK OF OPEN UNIVERSITY IN INDIA:

To provide library and information services to the distance learners, open Universities have a network of libraries. Indira Gandhi National Open University has two distinct categories of libraries.

(a) The central library as the apex at the University headquarters and,

(b) The Regional and Study Centres Ubraries as the branch libraries in the network.

The process of selection and acquisition, the collection and its organisation and the clientele vary a great deal from each other.

The Central library and Documentarion Centre is the combination of the features of academic as well as special libraries. The selection of library materials is the joint responsibility of the Librarian and academic staff. Books, journals, non-book materials such as films, audio-video cassettes, slides, microfilms and microfiches, maps, charts, pictures, globe and models are acquired to help the acap.ernic staff in the production of course materials for print and nonrprint media. The organisation of the collection is proposed to be fully automated. All the house-keeping operations of the library are to be computerised and the network is to be established with the Regional and Study Centre Libraries. The central library caters to the needs of academic and administrative staff of the main university. The editors, writers, education technologists and experts selected to work on part-time basis for the preparation of course materials are also entitled to use the central library resources.

Acquisition being a centralised process at the Open University Library, multiple copies of the books have to be acquired for the Regional and Sttidy Centre Libraries. These books are selected by the academic staff of the various disciplines keeping in mind the availability and readability of the books. The books are selected according to the course requirements and standards of the students. As far as possible, books recommended by the course writers as "Suggested readings" at the end of each unit (or block) are required for Regional and Study Centre Libraries. The books are processed ie., accessioned, classified and catalogued at the Central Library. The publishers/wholesalers are asked to mail the parcels of books directly to the Regional and Study Centres' libraries with two receipts as per the lists supplied and request the coordinator to send one to the centrallibra.ry and the other to the supplier. The invoices are sent to the Central Library. The books are accessioned and the bills are passed for the payment after receiving the receipts from the coordinators of the Regional and Study Centres. As the accession numbers of each book vary from one centre to other, the lists of books along with their accession numbers are sent to Regional and Study Centres. The computerised acquisition from catalogue print outs will be sent as soon as they are ready. The users of the Regional and Study Centre' libraries are the students, the part-time counsellors

appointed to impart guidance to the learners in each subject, the coordinator and his supporting staff.

2.6 CONCLUSION:

The University libraries in India are well equipped and organised as a result of substantial grants provided by the University Grants Commission but they are limited in number and are situated only in the big cities. Thus, they can cater only to the small percentage of urban students but cannot serve motivated learners in the rural and remote areas. The special libraries having specialised collection of advanced level in one or two disciplines are not immediately useful for all the students of liberal arts, undergraduate and professional courses. Hence, Open Universities should establish inter library loan arrangement with the University and the Special Libraries so as to borrow the books required by their research scholars and academic staff engaged in the preparation of courses. At this juncture, we can turn only to the public libraries to serve the library and information needs of the potential group of distance learners.

CHAPTER 3 : WEB DESIGNING

- WEB DESIGNING TECHNOLOGIES
- WEB DESIGNING STANDARDS
- WEB DESIGNING TECHNIQUES
- WEB DESIGNING TOOLS

3.1 INTERNET TECHNOLOGY:

3.11 Introduction:

During the past two decades, the world has witnessed a technological evolution that has provided a medium of communications entirely new to mankind. Through the use of networks, information in all forms has been disseminated throughout the world. This is known today as the World Wide Web (WWW) grew out of a project that began with a different intent (ARPANET). The ARPANET was designed and developed in 1969 by Bolt, Beranek and Newman under a contract for the Advanced Research Project Agency (ARPA) of the US Department of Defence. The purpose of the Network was to study how researchers could share data and how communications could be maintained in the event of a nuclear attack. The ARPANET Project was eventually turned over to the National Science Foundation (NSF) and ultimately became known as "Internet", by which the NSF allowed access to businesses, universities and individuals. In the beginning, many resources such as electronic mail, news, telnet, FTP, and Gopher were offered through the Internet to its users.

One of the early applications of the Internet was its most popular application, the World Wide Web (WWW) or sometimes known as "the Web". The WWW is one of the software tools that through the use of hypertext allows computers to link information in new ways, different from a sequential reading approach, to make it easy to retrieve and add information from different computer sources through the use of communication links (Berbers-Lee *et al.,* 1992). In the short time since its inception, the Internet has indeed revolutionized business, in that it redefines the methods used in traditional business practices and offers another important channel for mass communication (Foo and Lim, 1997).

During the early days of the Internet, the technology was primarily utilized as a medium for communication (e.g. e-mail) purposes. Soon afterwards many organizations from both the public and the private sectors began to discover that, in addition to use of the Internet and its popular WWW, they could utilize this technology in support of marketing and information dissemination purposes. This resulted in companies realizing that the greatest payback in investing in the technologies of WWW would be in sharing information

about the firms' products and services to the firms' stakeholders (Gardner, 1997). As a result, successful organizations of all sizes and types have been adopting different applications/technologies of WWW in discovering emerging ways of doing business that even a decade ago could not be imagined (Prawitt *et al.*, 1997). In recent years, the WWW has become the glittering palace of information and electronic trading that some visionary pundits promised (Jacobs, 1998). The Web has provided many improvements in the marketing business sector, particularly in areas such as "identification of sales prospects", "immediate access to information (i.e. product/service specifications and pricing) and allowing customers to obtain goods regardless of their geographical locations around the world (Hacker, 1996; Presti, 1996).

According to Bird (1996) the main reasons why businesses are utilizing the Web are primarily marketing related. They use the Web to:

- establish a presence;
- network;
- make business information available
- serve customers;
- heighten public interest;
- release time-sensitive data;
- sell products and services;
- reach a highly desirable demographic market;
- answer frequently asked questions;
- stay in contact with salespeople;
- open international markets;
- create a 24-hour service;
- make changing information available quickly;
- allow feedback from customers;
- test-market new services and products;
- reach the media; and
- reach a specialized market.

The data illustrate the changing foundation of marketing based on the emergence of the Web. Although the financial marketing advantages are not yet proven, the Web remains a fairly inexpensive form of communicating with potential customers (Bird, 1996).

In addition to the use of the Web for communication and marketing purposes, during the past two decades, there have been many other emerging Web-enabled technologies, including:

- mail technologies;
- electronic interchange;
- electronic data interchange (EDI);
- electronic commerce (EC);
- network management;
- organizational intranets/extranets;
- online analytical processing (OLAP); and
- teleconferencing, etc.

3.2 WORLD WIDE WEB CONSORTIUM (W3C):

The W3C was created to lead the Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. It is an international industry consortium jointly run by the MIT Laboratory for Computer Science (MIT LCS) in the USA, the National Institute for Research in Computer Science and Control (INRIA) in France and Keio University in Japan. Services provided by the Consortium include:

- A repository of information about the World Wide Web for developers and users,
- Reference code implementations to embody and promote standards,and
- Various prototype and sample applications to demonstrate use of new technology.

At present over 500 organizations are Members of the Consortium.

3.3 WEB DESIGNING TECHNOLOGIES

3.31 HYPERTEXT MARKUP LANGUAGE (HTML)

HTML is the *lingua franca* for publishing hypertext on the World Wide Web. It is a non-proprietary format based upon SGML, and can be created and processed by a wide range of tools, from simple plain text editors - you type it in from scratch to sophisticated WYSIWYG authoring tools. HTML uses tags such as <h1> and </h1> to structure text into headings, paragraphs, lists, hypertext links etc.

3.311 EXTENSIBLE HYPERTEXT MARKUP LANGUAGE (XHTML):

The Extensible HyperText Markup Language (XHTML) is a family of current and future document types and modules that reproduce, subset, and extend HTML, reformulated in XML. XHTML Family document types are all XML-based, and ultimately are designed to work in conjunction with XML-based user agents. XHTML is the successor of HTML, and a series of specifications has been developed for XHTML.

3.312 HTML Working Group:

To develop the next generation of HTML as a suite of XML tag sets with a clean migration path from HTML 4. Some of the expected benefits include: reduced authoring costs, an improved match to database & workflow applications, a modular solution to the increasingly disparate capabilities of browsers, and the ability to cleanly integrate HTML with other XML applications. For further information, see the Charter <./2002/05/html/charter> for the HTML Working Group <Group/> (*members only <http://cgi.w3.org/MemberAccess/>*).The HTML Working Group Charter has been renewed in August 2002.

3.313 W3C Recommendations:

W3C produces what are known as Recommendation. These are specifications, developed by W3C working groups, and then reviewed by Members of the Consortium. A

W3C Recommendation indicates that consensus has been reached among the Consortium Members that a specification is appropriate for widespread use.

XHTML 1.0 is the W3C's first Recommendation for XHTML, following on from earlier work on HTML 4.01, HTML 4.0, HTML 3.2 and HTML 2.0. With a wealth of features, XHTML 1.0 is a reformulation of HTML 4.01 in XML, and combines the strength of HTML 4 with the power of XML.

XHTML 1.0 is the first major change to HTML since HTML 4.0 was released in 1997. It brings the rigor of XML to Web pages and is the keystone in W3C's work to create standards that provide richer Web pages on an ever increasing range of browser platforms including cell phones, televisions, cars, wallet sized wireless communicators, kiosks, and desktops.

XHTML 1.0 is the first step and the HTML Working Group is busy on the next. XHTML 1.0 reformulates HTML as an XML application. This makes it easier to process and easier to maintain. XHTML 1.0 borrows elements and attributes from W3C's earlier work on HTML 4, and can be interpreted by existing browsers, by following a few simple guidelines. This allows to start using XHTML. One can roll over your old HTML documents into XHTML using an Open Source HTML Tidy utility. This tool also cleans up markup errors, removes clutter and prettifies the markup making it easier to maintain.

3.314 Three "flavors" of XHTML 1.0:

XHTML 1.0 is specified in three "flavors". You specify which of these variants you are using by inserting a line at the beginning of the document. For example, the HTML for this document starts with a line, which says that it is using XHTML, 1.0 Strict. Thus, if you want to validate the document, the tool used knows which variant you are using. Each variant has its own DTD - Document Type Definition - that sets out the rules and regulations for using HTML in a succinct and definitive manner.

3.3141 XHTML 1.0 Strict- Use this when you want really clean structural mark-up, free of any markup associated with layout. Use this together with W3C's Cascading Style Sheet language (CSS) to get the font, color, and layout effects you want.

3.3142 XHTML 1.0 Transitional- Many people writing Web pages for the general public to access might want to use this flavor of XHTML 1.0. The idea is to take advantage of XHTML features including style sheets but nonetheless to make small adjustments to your markup for the benefit of those viewing your pages with older browsers which can't understand style sheets. These include using the body element with bgcolor, text and link attributes.

3.3143 XHTML 1.0 Frameset- Use this when you want to use Frames to partition the browser window into two or more frames.

The complete XHTML 1.0 specification is available in English in several formats, including HTML, PostScript and PDF.

3.315 HTML 4.01:

HTML 4.01 is a revision of the HTML 4.0 Recommendation first released on 18th December 1997. The revision fixes minor errors that have been found since then. The XHTML 1.0 spec relies on HTML 4.01 for the meanings of XHTML elements and attributes. This allowed us to reduce the size of the XHTML 1.0 spec very considerably.

3.316 XHTML Basic:

XHTML Basic is the second Recommendation in a series of XHTML specifications. The XHTML Basic document type includes the minimal set of modules required to be an XHTML Host Language document type, and in addition it includes images, forms, basic tables, and object support. It is designed for Web clients that do not support the full set of XHTML features; for example, Web clients such as mobile phones, PDAs, pagers, and setup boxes. The document type is rich enough for content authoring.

XHTML Basic is designed as a common base that may be extended. For example, an event module that is more generic than the traditional HTML 4 event system could be added or it could be extended by additional modules from XHTML Modularization such as the Scripting Module. The goal of XHTML Basic is to serve as a common language supported by various kinds of user agents.

3.3161 Modularization of XHTML:

Modularization of XHTML is the third Recommendation in a series of XHTML specifications. This Recommendation specifies an abstract modularization of XHTML and an implementation of the abstraction using XML Document Type Definitions (DTDs). This modularization provides a means for subsetting and extending XHTML, a feature needed for extending XHTML's reach onto emerging platforms.

Modularization of XHTML will make it easier to combine with markup tags for things like vector graphics, multimedia, math, electronic commerce and more. Content providers will find it easier to produce content for a wide range of platforms, with better assurances as to how the content is rendered.

The modular design reflects the realization that a one-size-fits-all approach will no longer work in a world where browsers vary enormously in their capabilities. A browser in a cell phone can't offer the same experience as a top of the range multimedia desktop machine. The cell phone doesn't even have the memory to load the page designed for the desktop browser.

3.3162 XHTML 1.1 - Module-based XHTML:

This Recommendation defines a new XHTML document type that is based upon the module framework and modules defined in Modularization of XHTML. The purpose of this document type is to serve as the basis for future extended XHTML 'family' document types, and to provide a consistent, forward-looking document type cleanly separated from the

deprecated, legacy functionality of HTML 4 that was brought forward into the XHTML 1.0 document types.

This document type is essentially a reformulation of XHTML 1.0 Strict using XHTML Modules. This means that many facilities available in other XHTML Family document types (e.g., XHTML Frames) are not available in this document type. These other facilities are available through modules defined in Modularization of XHTML, and document authors are free to define document types based upon XHTML 1.1 that use these facilities (see Modularization of XHTML for information on creating new document types).

3.3163 Difference between XHTML1.0, XHTML Basic and XHTML1.1:

The first step was to reformulate HTML 4 in XML, resulting in XHTML 1.0. By following the HTML Compatibility Guidelines <http://www.w3.org/TR/xhtml1/> set forth in Appendix C of the XHTML 1.0 specification, XHTML 1.0 documents could be compatible with existing HTML user agents.

The next step is to modularize the elements and attributes into convenient collections for use in documents that combine XHTML with other tag sets. The modules are defined in Modularization of XHTML. XHTML Basic is an example of fairly minimal build of these modules and is targeted at mobile applications.

XHTML 1.1 is an example of a larger build of the modules, avoiding many of the presentation features. While XHTML 1.1 looks very similar to XHTML 1.0 Strict, it is designed to serve as the basis for future extended XHTML Family document types, and its modular design makes it easier to add other modules as needed or integrate itself into other markup languages. XHTML 1.1 plus MathML 2.0 http://www.w3.org/TR/MathML2/appendixa.html> document type is an example of such XHTML Family document type.

3.317 Previous Versions of HTML:

3.3171 HTML 4.0:

First released as a W3C Recommendation on 18 December 1997. A second release was issued on 24 April 1998 with changes limited to editorial corrections. This specification has now been superseded by HTML 4.01.

3.3172 HTML 3.2:

W3C's first Recommendation for HTML, which represented the consensus on HTML features for 1996. HTML 3.2 added widely-deployed features such as tables, applets, text-flow around images, superscripts and subscripts, while providing backwards compatibility with the existing HTML 2.0 Standard.

3.3173 HTML 2.0:

HTML 2.0 (RFC 1866 <http://www.rfc-editor.org/rfc/rfc1866.txt>) was developed by the IETF's HTML Working Group, which closed in 1996. It set the standard for core HTML features based upon current practice in 1994. Note that with the release of RFC 2854 <http://www.rfc-editor.org/rfc/rfc2854.txt>, RFC 1866 has been obsolete and its current status <http://www.ietf.org/iesg/1rfc_index.txt> is historic.

3.3174 ISO HTML:

ISO/IEC 15445:2000 <http://purl.org/NET/ISO+IEC.15445/15445.html> is a subset of HTML 4, standardized by ISO/IEC. It takes a more rigorous stance for instance, an h3 element can't occur after an h1 element unless there is an intervening h2 element.

3.318 Modularization of XHTML in XML Schema:

The purpose of this document is to describe a modularization framework for languages within the XHTML Namespace using XML Schema. This document provides a complete set of XML Schema modules for XHTML. In addition to the schema modules themselves, the framework presented here describes a means of further extending and modifying XHTML.

3.319 XML Events (This specification was renamed from "XHTML Events".):

The XML Events module defined in this specification provides XML languages with the ability to uniformly integrate event listeners and associated event handlers with Document Object Model (DOM) Level 2 event interfaces. The result is to provide an interoperable way of associating behaviors with document-level markup.

3.3191 An XHTML + MathML + SVG Profile:

A XHTML+MathML+SVG profile is a profile that combines XHTML 1.1, MathML 2.0 and SVG 1.1 together. This profile enables mixing XHTML, MathML and SVG in the same document using XML namespaces mechanism, while allowing validation of such a mixed-namespace document. This specification is a joint work with the SVG Working Group, with the help from the Math Working Group.

3.3192 XHTML 2.0:

XHTML 2.0 is a markup language intended for rich, portable web-based applications. While the ancestry of XHTML 2.0 comes from HTML 4, XHTML 1.0, and XHTML 1.1, it is *not* intended to be backward compatible with its earlier versions. Application developers familiar with earlier its ancestors will be comfortable working with XHTML 2.0.

XHTML 2 is a member of the XHTML Family of markup languages. It is an XHTML Host Language as defined in Modularization of XHTML. As such, it is made up of a set of XHTML Modules that together describe the elements and attributes of the language, and their content model. XHTML 2.0 updates many of the modules defined in Modularization of XHTML, and includes the updated versions of all those modules and their semantics. XHTML 2.0 also uses modules from Ruby, XML Events, and XForms.

3.3193 Xframes:

XFrames is an XML application for composing documents together, replacing HTML Frames. XFrames is *not* a part of XHTML per se, which allows similar functionality to

HTML Frames, with fewer usability problems, principally by making the content of the frameset visible in its URI.

3.3194 XHTML 1.0 in XML Schema:

This document describes *informative* XML Schemas for XHTML 1.0. These Schemas are still work in progress, and this document *does not* change the normative definition of XHTML 1.0.

3.3195 HLink:

The HLink module defined in this specification provides XHTML Family Members with the ability to specify which attributes of elements represent Hyperlinks, and how those hyperlinks should be traversed, and extends XLink use to a wider class of languages than those restricted to the syntactic style allowed by XLink.

3.3196 GUIDELINES FOR AUTHORING:

Some of the guidelines for authoring an HTML documents are discussed here. These are essential to end up with pages that are easy to maintain, look acceptable to users regardless of the browser they are using, and can be accessed by the many Web users with disabilities. Meanwhile W3C have produced some more formal guidelines for authors.

3.31961 Nature of style sheets:

For most people the look of a document - the color, the font, the margins - are as important as the textual content of the document itself. But make no mistake. HTML is not designed to be used to control these aspects of document layout. What you should do is to use HTML to mark up headings, paragraphs, lists, hypertext links, and other structural parts of your document, and then add a style sheet to specify layout separately, just as you might do in a conventional Desk Top Publishing Package. That way, not only is there a better chance of all browsers displaying your document properly, but also, if you want to change such things as the font or color, it's really simple to do so.

3.31962 FONT tag considered harmful:

Many filters from word-processing packages, and also some HTML authoring tools, generate HTML code, which is completely contrary to the design goals of the language. What they do is to look at a document almost purely from the point of view of layout, and then mimic that layout in HTML by doing tricks with FONT, BR and & nbsp; (non-breaking spaces). HTML documents are supposed to be structured around items such as paragraphs, headings and lists. Yet some of these documents barely have a paragraph tag in sight.

The problem comes when the content of pages needs to be updated, or given a new layout, or re-cast in XML (which is now to be the new mark-up language). With proper use of HTML, such operations are not difficult, but with a muddle of non-structural tags it's quite a different matter; maintenance tasks become impractical. To correct pages suffering from injudicious use of FONT, try the HTML Tidy program, which will do its best to put things right and generate better and more manageable HTML.

3.31963 Make your pages readable by those with disabilities:

The Web is a tremendously useful tool for the visually impaired or blind user, but bear in mind that these users rely on speech synthesizers or Braille readers to render the text. Sloppy mark-up, or mark-up which doesn't have the layout defined in a separate style sheet, is hard for such software to deal with. Wherever possible, use a style sheet for the presentational aspects of your pages, using HTML purely for structural mark-up.

It also include descriptions with each image, and try to avoid server-side image maps. For tables, you should include a summary of the table's structure, and remember to associate table data with relevant headers. This will give non-visual browsers a chance to help orientates people as they move from one cell to the next. For forms, remember to include labels for form fields.

3.3197 W3C Markup Validation Service:

To further promote the reliability and fidelity of communications on the Web, W3C has introduced the W3C Markup Validation Service <http://validator.w3.org/> at http://validator.w3.org/. Content providers can use this service to validate their Web pages against the HTML and XHTML Recommendations, thereby ensuring the maximum possible audience for their Web pages. It also supports XHTML Family document types such as XHTML+MathML and XHTML+MathML+SVG, and also other markup vocabularies such as SVG <../Graphics/SVG/>.

Software developers who write HTML and XHTML editing tools can ensure interoperability with other Web software by verifying that the output of their tool complies with the W3C Recommendations for HTML and XHTML.

3.3198 HTML Tidy:

HTML Tidy is a stand-alone tool for checking and pretty-printing HTML that is in many cases able to fix up mark-up errors, and also offers a means to convert existing HTML content into well-formed XML, for delivery as XHTML. Dave Raggett originally wrote HTML Tidy, and it is now maintained as an open source project at Source Forge <http://tidy.sourceforge.net/> by a group of volunteers.

3.31991 MAINTENANCE OF HTML/XHTML PAGE:

While editing HTML it's easy to make mistakes. There must be some provision to sort out these mistakes automatically and tidy up sloppy editing into nicely laid out markup. HTML TIDY is a free utility originally developed by Dave Raggett's for providing the same facility. It also works great on the atrociously hard to read markup generated by specialized HTML editors and conversion tools, and can help in identifying where you need to pay further attention on making your pages more accessible to people with disabilities.

Tidy is able to fix up a wide range of problems and to bring to your attention things that you need to work on yourself. Each item found is listed with the line number and column

so that you can see where the problem lies in your markup. Tidy won't generate a cleaned up version when there are problems that it can't be sure of how to handle. These are logged as "errors" rather than "warnings".

At present Tidy is being maintained by a group of volunteers working together as part of the open source community at Source Forge. The source code continues to be available under an open source license.

3.31992 Internationalization issues:

Tidy offers you a choice of character encoding as US ASCII, ISO Latin-1, UTF-8 and the ISO 2022 family of 7 bit encoding. The full set of HTML 4.0 entities is defined. Cleaned up output uses HTML entity names for characters when appropriate. Otherwise characters outside the normal range are output as numeric character entities.

3.32 XML LINKING:

3.321 Introduction:

XML Linking Language (XLink) allows elements to be inserted into XML documents in order to create and describe links between resources. It uses XML syntax to create structures that can describe the simple unidirectional hyperlinks of today's HTML, as well as more sophisticated links.

3.322 Tools:

- X2X from empolis UK Ltd. is an XML XLink Engine. X2X allows linking between documents and information resources without needing to change the resources that are being linked. X2X removes the requirement to insert link information inside document content. The Links are NOT in the document.

- Fujitsu XLink Processor: Fujitsu XLink Processor, which is developed by Fujitsu Laboratories Ltd., is an implementation of XLink and XPointer.

- xlinkit.com: is a lightweight application service which provides rule-based XLink generation and checks the consistency of distributed documents and web content. You tell xlinkit.com the information you want to link and rules that relate the information. xlinkit.com will generate the links that you can then use for navigation. It will also diagnose inconsistent information.

- Mozilla: The Open source browser has support for XLinks simple links.

- Amaya: The W3C editor/browser now supports XLinks simple links too.

- XTooX is a free XLink processor that turns extended out-of-line links into inline links. It takes as its input a link base- a document containing only XLinks - and puts the links into the referenced documents. XTooX is available under the GNU Lesser General Public License.

3.33 XML BASE:

This specification proposes syntax for providing the equivalent of HTML BASE functionality generically in XML documents by defining an XML attribute named xml base.

3.34 XML POINTER LANGUAGE (XPointer):

3.341 Introduction:

XML Pointer Language (XPointer) is the language to be used as a fragment identifier for any URI-reference that locates a resource of Internet media type. XPointer has been split into a framework for specifying location schemes, and three schemes: element(), xmlns() and xpointer(). The framework and the first two schemes form the XPointer Recommendation, and provide a minimal inventory of mechanisms.

The xpointer() scheme, which is based on the XML Path Language (XPath), is still under development. It supports addressing into the internal structures of XML documents. It allows for traversals of a document tree and choice of its internal parts based on various properties, such as element types, attribute values, character content, and relative position.

3.342 Tools:

- Fujitsu XLink Processor: Fujitsu XLink Processor, which is developed by Fujitsu Laboratories Ltd., is an implementation of XLink and (almost all of) XPointer.

- libxml: the Gnome XML library has a beta implementation of XPointer. The full syntax is supported but the test suite does not cover all aspects yet.

- 4XPointer: this is an XPointer Processor Written in Python by Fourthought, Inc <http://www.fourthought.com/>.

- At the University of Bologna two different implementations of XPointer are in progress, one in JavaScript for ASP pages and another in Java.

- XPointerLib from the Connexions project, a mozdev.org project providing XPointer support for Mozilla / Netscape 7 / Phoenix browsers. It is an XPCOM service written in JavaScript that creates and resolves a subset of the XPointer language.

3.35 MATHEMATICAL MARKUP LANGUAGE (MathML):

3.351 Introduction:

The World Wide Web Consortium (W3C) has issued its first XML- based applications as a recommendation. The Mathematical Markup Language (MathML) is a way of describing math syntax so mathematical ideas can be exchanged using the web. A W3C recommendation means the consortium considers the specification to be stable makes a

contribution to web interoperability and has been thoroughly reviewed by the W3C membership. The W3C is not a standards body, it is merely a place where ideas can be discussed and a recommendation is far as a spec can go within the organization. The MathML spec provides for two sets of markup tags: a set that presents the mathematical notations and a set that relays the semantic meaning of expressions. MathML is not intended as user language; it is meant for use with software tools that translate mathematical equations into a human-readable format. The consortium says such tools are already under development, both as freeware and as commercial products. XML, or eXtensible Markup Language, is a version of the ISO- standardized Standard Generalized Markup Language (SGML), and is viewed as the successor to HTML. It provides for more wide-ranging applications than merely displaying text and pictures. For instance, the sets of tags are extensible, as the name suggests, so developers can make up their own tags as required, which can be used to identify a particular piece of text as belonging to a particular group. It also retains some of SGML's basic features, such as complex structures, validation and human readability.

MathML is an XML-compliant markup language that has two tag sets. The first describes the notation of mathematical data; the second contains the semantic meaning of mathematical expressions. Users who want to publish a mathematical equation on the Web must render and display the equation as an image. MathML shares XML's capability to retrieve data objects in real time from external databases and monitors being used as retrievable computer entities. One analyst said the slide-rule community will initially rejoice, but the implications of MathML reach further.

According to Martin Marshall, an industry analyst at Zona Research, in Redwood City, Calif "This is for the engineering and scientific community what HTML was to the rest of us. Even without the scientific notation, it allows for a lot of secondary calls."

MathML 1.0 provides a solid foundation for representing mathematical expressions. However, a number of critical requirements dating back to the original HTML-Math Working Group Charter remain to be accomplished, and other goals developed as a result of

MathML implementations or of feedback from the community remain to be met. MathML 1.0 offered a unique opportunity to ensure effective math on the Web through its widespread acceptance, and it seems very desirable to maintain the present MathML 1.0 Recommendation, and to further develop the specification.

The current Math Working Group proposes that the W3C establish a new Math Working Group to continue the work of the W3C Math activity. The proposed revision MathML aims to reduce the overhead involved in publishing scientific and technical Web content, while increasing its scope to accommodate new areas of science. We expect that as a result of a new MathML the suite of tools for authoring, managing, transforming and rendering MathML will continue to evolve and leverage the relationship between MathML and other W3C specifications.

On 21st February, 2001 World Wide Web Consortium (W3C) released of the Mathematical Markup Language (MathML) 2.0 as a W3C Recommendation. MathML 2.0, an XML application, provides encoding mathematical notation and content for use on the Web. A W3C Recommendation indicates that a specification is stable, contributes to Web interoperability, and has been reviewed by the W3C Membership, who are in favor of supporting its adoption by academic, industry, and research communities. MathML 2.0 Extends the Foundation for Math on the Web. MathML 2.0 consist of a number of XML tags that can be used to markup an equation in terms of its presentation and also its semantics. As a result, MathML 2.0 attempts to capture something of the meaning behind equations rather than concentrating entirely on how they are going to be formatted out on the screen. This is because mathematical equations are meaningful to many applications independent of how they are rendered aurally or visually.

According to Vincent Quint, W3C User Interface Domain Leader "What HTML did for text on the Web? MathML 2.0 does for the language of mathematics, and because it is written in XML, it makes it possible for Math content to be not only displayed, but able to be reused and transformed by other applications on the Web." MathML 2.0 is intended to facilitate the use and re-use of mathematical and scientific content on the Web, and for other

applications such as computer algebra systems, print typesetting, and voice synthesizers. MathML can be used to encode both the presentation of mathematical notation for high-quality visual display, and mathematical content, for applications where the semantics plays more of a key role such as scientific software or voice synthesis.

3.352 MathML 2.0 Integrates W3C Technologies:

MathML 2.0 builds on MathML 1.0 by extending the set of symbols and expressions and through improved integration of other W3C technologies. Users of MathML 2.0 are now able to combine it with other W3C technologies to make more dynamic and varied content:

- Equations can be styled with Cascading Style Sheets (CSS),
- Links can be associated to any math expression through XML Linking Language (XLink), and
- MathML elements can be seamlessly included in XHTML documents with namespaces.
- MathML 2 also includes the MathML Document Object Model (MathML DOM), which provides a more convenient, and MathML-specific way to identify MathML components and enable any scripting language to manipulate it.

The Math Working Group has produced test suites, and is already at work developing an XML Schema for MathML 2, as well as a hybrid schema to combine XHTML and MathML 2.0.

3.353 MathML and Technologies on various Operating Systems:

MathML in Web pages makes it possible to be viewed on a large number of browsers. It also configures browsers to make them able to display MathML. Currently the browsers that will render the pages using the conventions below are:

3.3531 Windows:

- IE 5.0 with the Techexplorer plug-in
- IE 5.5 with either the MathPlayer or Techexplorer plug-ins
- IE 6.0, optionally with MathPlayer or Techexplorer plug-ins
- Netscape 6.1 with Techexplorer plug-in
- Netscape 7.0 PR1
- Amaya (Presentation MathML only)
- Mozilla 0.9.9

3.3532 Macintosh:

- IE 5.0 with Techexplorer plug-in
- Mozilla 0.9.9

3.3533 Linux/Unix:

- Netscape 6.1 with Techexplorer plug-in
- Netscape 7.0 PR1
- Mozilla 0.9.9
- Amaya (Presentation MathML only)

3.36 SYNCHRONIZED MULTIMEDIA INTEGRATION LANGUAGE (SMIL):

3.361 Introduction:

The World-Wide Web has grown up in an ad hoc way, starting with text, then adding images, sounds and video. So the simultaneous use of these multimedia elements has never been addressed properly. In particular, the only way to create a constantly changing flux of text, sounds and images is to create a video stream. This approach is both inflexible and inefficient, as video tends to require high bandwidth.

The Synchronized Multimedia Integration Language (SMIL, pronounced "smile") enables simple authoring of interactive audiovisual presentations. SMIL is typically used for "rich media"/multimedia presentations which integrate streaming audio and video with images, text or any other media type. SMIL is an easy-to-learn HTML-like language, and many SMIL presentations are written using a simple text-editor. SMIL permits multimedia streams to be played sequentially or in parallel, and for different elements to be placed in absolute positions on the screen. Its hotlinks can be embedded in video multimedia elements, as well as in text and images. This would allow SMIL presentations to offer full interactivity.

SMIL enables authors to bring television-like content to the Web, avoiding the limitations of traditional television and significantly lowering the bandwidth requirements for transmitting this type of content over the Internet. With SMIL, producing audio-visual content does not require learning a programming language and can be done using a simple text editor. SMIL was developed by the W3C Synchronized Multimedia (SYMM) Working Group, a mix of experts from the four divergent industries (CD-ROM), interactive television, Web, and audio/video streaming) interested in bringing synchronized multimedia to the Web. Philipp Hoschka, chairman of the W3C SYMM Working Group and editor of the SMIL specification says "Such an agreement is the necessary signal for content providers to start creating synchronized multimedia content for the Web, and, thus, a prerequisite for market growth in this area."

3.362 Features:

SMIL offers the following key features:

- Easy-to-learn synchronization primitives: 90 percent of the power of SMIL can be tapped by mastering only two tags, "parallel" and "sequential."

- Temporal hyperlinking: This feature offers all the capabilities of hyperlinks in HTML and adds capabilities required in time-based presentations.

- Reusability of media objects: All components of the multimedia presentation are referenced via URLs rather than physically embedded into a SMIL file. For example, videos stored in a digital video library can be reused in many presentations.

- Load balancing: Different media objects in a presentation can be stored on different servers--another benefit of using URLs rather than physically including media objects within the SMIL document.

- Language selection: Authors can indicate that an audio track is available in several languages, thus increasing the potential audience of the content.

- Bandwidth selection: Authors can express that a media object such as an audio track is available in different versions, each having been encoded for a different transmission bandwidth. This guarantees that presentations can be played even when only low-bandwidth access is available.

To understand the utility of SMIL we can analyze television news broadcast as an example, large parts of the screen contain text, still images, and graphical elements with full-motion video occupying only a small part of the screen real estate. These media types can all be included on a Web page today. However, the Web lacks a simple way to express synchronization over time such as "play audio file A in parallel with video file B" or "show image C after audio file A has finished playing." SMIL enables this type of information to be expressed easily, allowing television-like content to be created on the Web.

SMIL avoids having to swamp the Internet with high-bandwidth video if you want to create television-like content," said Hoschka. When rewritten in SMIL, many of today's television news broadcasts would require far less bandwidth, eliminating the need to convert low-bandwidth media types such as text and images into high-bandwidth video. Till 1998 RealNetworks and Allaire Corp. have been the only major vendors to announce significant SMIL-supporting products.

3.363 SMIL: Prime Versions:

3.3631 SMIL 1.0:

SMIL 1.0 enables authors to bring TV-like content to the Web, avoiding the limitations of traditional television and lowering the required Internet bandwidth for this type of content. With SMIL, producing audio-visual presentations for the Web is easy, since it can be done using a simple text editor, and does not require learning a programming language.

SMIL 1.0 is the W3C Recommendation (standard) for Web-based multimedia first implemented by RealNetworks with the advent of RealSystem G2 in June of 1998. RealNetworks co-authored the SMIL 1.0 specification. SMIL 1.0 enables the delivery of long format, Web-based multimedia to a broad range of audiences, from modems to T3 Internet connections. As an open XML-based language, SMIL enables a wide range of audio-visual presentation authoring environments, ranging from simple text editors to graphical editing tools such as RealNetworks RealSlideshow.

3.3632 SMIL Boston:

Next version of XML-based multimedia language features reusable modules, generic animation, improved interactivity and TV integration Leading the Web to its full potential. SMIL Boston builds upon the W3C SMIL 1.0 Recommendation, and adds important extensions, including reusable modules, generic animation, improved interactivity, and TV integration, all written in the Extensible Markup Language (XML).

The SMIL Boston Working Draft proposes several extensions to SMIL 1.0, such as integration with TV broadcasts, animation functionality, improved support for navigation of timed presentations, and the ability to integrate SMIL markup in other XML-based languages. These extensions are based on the feedback received from authors, implementers and others using the SMIL 1.0 infrastructure existing today.

SMIL Boston Modules Enable Integration with other XML-based Languages Designing the syntax and semantics of a markup language requires significant time and

effort. Fortunately, designers of other XML-based languages are able take full advantage of SMIL Boston, as it is designed as a set of reusable modules. With SMIL Boston, language designers can for example add timing information to Extensible HyperText Markup Language (XHTML) and Scalable Vector Graphics (SVG), simply by importing the SMIL Boston Timing and Synchronization module, rather than building timing models and syntax from scratch.

SMIL Boston Enables Creation of Animations in XML Animation is a popular approach to create compelling Web content while reducing the download time for a presentation. While the most popular form of animation on the Web today is animated GIF, it has several limitations. As the animation is encoded in binary format, one needs special editing tools to create it. Further, only GIF images can be used in the animation- one cannot include a JPEG image, or an XHTML headline, or an SVG vector graphics object.

The SMIL Boston animation module eliminates the limitations found of the animated GIF format. Since SMIL Boston modulesnare based on XML, animations can be written using a simple text editor. It enables animation of any media format, such as JPEG images, PNG images, even video clips. The SMIL Boston animation module can also be used to add animation capabilities to other XML-based languages, such as XHTML, SVG or an XML-based 3D language.

One of the benefits of SMIL presentations over traditional TV content is that users can navigate within the presentation, thereby focusing on the parts of the presentation that interests them most. This can be achieved by providing a table of contents of the presentation.

Using SMIL Boston, the table of contents and the content itself can be contained in the same SMIL file, rather than being split over several files. This simplifies authoring, and reduces delays when users navigate through the presentation.

Another benefit over traditional TV content is that SMIL allows authors to include additional content (e.g. background information) on the topic of the presentation. In SMIL

Boston, optional parts can be contained in the same SMIL file as the main presentation. This allows the user to access optional content without interrupting the main presentation.

Future digital television broadcasts are to use very similar techniques as today's SMIL presentations. Rather than broadcasting audio and video signals only, digital TV broadcasts may consist of a combination of images, text and other media objects that are synchronized at the receiver. Some of the capabilities of SMIL Bostan are:

- New Transitions module describing transitions within SMIL and other XML-based documents
- Improved control for runtime content choices
- New Metadata module to better describe SMIL documents published on the Web
- Improved interactivity
- Improved hyper navigation
- Improved integration into other XML-based languages
- Tighter integration with multimedia protocols such as RTP and RTSP

SMIL Boston modularizes SMIL functionality, providing standards based integration of SMIL functionality with other XML based languages and applications. Content authors and application developers both benefit from this flexibility: application developers can integrate needed functionality while content authors are able to build on their existing knowledge base.

3.3633 SMIL 2.0:

SMIL has grown substantially from its first incarnation. SMIL 2.0 describes modules for timing, animation, layout, linking, media objects, transitions, and more. The basic media types supported are animation, audio, image, text, text stream, and video. Some tag names from the 1.0 release have been updated or deprecated to make SMIL 2.0 compatible with the document object model.

So why should anyone give a heap of beans about SMIL? SMIL makes for compelling content. Video presentations on the Web basically stink today. A two-inch video popup, for example, has little visual impact. Animation files, a la Macromedia Flash, on the other hand, can generate impact and motion. Integrate the two techniques within a single presentation and you are getting somewhere.

Both animation and video suffer from some key problems, however. Imagine internationalizing a video. You might have to dub the video in six languages and make it available in three bandwidth-optimized versions. Eighteen files just to support a video clip. If you want to add optional captions for the hearing impaired, you're up to m36 files.

With the SMIL approach, the video, audio, and text components can be treated as individual synchronized streams. The video object can self-select the appropriate bandwidth version. The language for the audio and text stream can be chosen dynamically from user preferences as the SMIL page is composed. Thirty-six files have now been reduced to no more than 15 simpler files. More importantly, if there is a problem with a caption, it can be fixed in a text editor without impacting the other files. Now, add some bandwidth-friendly vector animation in an underlying region, and you've got a fine looking presentation.

3.364 SPECIFICATIONS:

3.3641 SMIL 2.0

- W3C Recommendation: Synchronized Multimedia Integration Language (SMIL 2.0) <http://www.w3.org/TR/smil20>

- Translations </AudioVideo/SMIL/translations> of SMIL 2.0 (e.g. Korean <http://www.smilmedia.com/spec/spec1~7.htm>)

- W3C Note "XHTML+SMIL Profile" <http://www.w3.org/TR/XHTMLplusSMIL/>

- SMIL 2.0 Testsuite <http://www.w3.org/2001/SMIL20/testsuite/>

- Implementation Results from SMIL2.0 Test suite <http://www.w3.org/2001/05/23/SMIL-Implementation-result.html>

- For SMIL profile used in 3GPP MMS <http://www.mobilemms.com/mmsfaq.asp> (Multimedia Messaging Service) and Streaming Service, see 3GPP specifications (TS 26.140 <ftp://ftp.3gpp.org/specs/latest/Rel-5/26_series/26140-510.zip> defines MMS and TS 26.234 <ftp://ftp.3gpp.org/specs/latest/Rel-5/26_series/26234-510.zip>, Section B and Appendix B define the MMS SMIL profile)

3.3642 SMIL 1.0

- W3C Recommendation: Synchronized Multimedia Integration Language (SMIL) 1.0 Specification </TR/REC-smil>

- Translations </AudioVideo/SMIL/translations> of SMIL 1.0 (e.g. Chinese <http://lightning.prohosting.com/~qqiu/smil/trans/REC-smil-19980615-cn.html>, German <http://www.sunshine-company.de/w3c/REC-smil-19980615-DE.html>, Italian <http://www.w3c.cnr.it/office/traduzioni/REC-smil-it.html>, Japanese <http://www.doraneko.org/misc/smil10/smil10.html>, Korean <http://www.mentallink.com/resource/smil/smil10-kr.html>, Portuguese <http://www.utad.pt/~leonelm/w3ctranslations/smil/>)

- SMIL 1.0 Player Testcases <http://smil.nist.gov/Testcase.html> and SMIL Player Feature List <http://smil.nist.gov/Feature.html>

3.365 PLAYERS:

3.3651 SMIL 2.0

- RealOne Platform <http://www.realnetworks.com/solutions/ecosystem/realone.html?src=rnhmfs> by RealNetworks with full support for the SMIL 2.0 Language profile.

- GRiNS for SMIL-2.0 <http://www.oratrix.com/GRiNS/SMIL-2.0/> by Oratrix provides a SMIL 2.0 player which supports SMIL 2.0 syntax and semantics.

- SMIL Player by InterObject <http://www.inobject.com/mmplay.htm>. The player supports SMIL 2.0 Basic Profile.
  The player runs on PC with Windows NT/2000/XP and handheld devices with Pocket PC, such as Compaq iPAQ. Refer to product specifications

- Internet Explorer 6.0 <http://www.microsoft.com/windows/ie/preview/default.asp> by Microsoft includes implementation of XHTML+SMIL Profile <http://www.w3.org/TR/2001/WD-XHTMLplusSMIL-20010807/> Working Draft

- Internet Explorer 5.5 <http://www.microsoft.com/windows/ie/default.htm> by Microsoft supports many of the SMIL 2.0 draft modules including Timing and Synchronization, BasicAnimation, SplineAnimation, BasicMedia, MediaClipping, and BasicContentControl. See an introductory article about SMIL 2.0 support (called HTML+TIME 2.0 <http://msdn.microsoft.com/workshop/Author/behaviors/htmltime.asp>) in IE 5.5.

- NetFront v3.0 <http://k-tai.impress.co.jp/cda/article/news_toppage/13103.html> is a micro browser for PDA/mobile phone/information appliances. It claims to support HTML 4.01/XHTML 1.0/ SMIL Basic/SVG Tiny.

- Pocket SMIL <http://wam.inrialpes.fr/software/pocketsmil/>, it is written in C++.

- RubiC <http://www.roxia.co.kr> is developed by Roxia Co.,Ltd. It includes an authoring tool and player, and fully supports SMIL 2.0 specification. "RubiC" is also available for mobile handset for mobile internet MMS(Multimedia Messaging Service)

- List of MMS Simulators <http://lists.w3.org/Archives/Public/www-mobile/2002Aug/0007.html>

3.3652 SMIL 1.0

- Grins (SMIL1.0) <http://www.oratrix.com/GRiNS/index.html> by Oratrix

- HPAS <http://www.research.digital.com/SRC/HPAS> by Compaq

- Lp player <http://www.prodworks.com/> by Productivity Works

- QuickTime 4.1 <http://www.apple.com/quicktime/authoring/qtsmil.html> by Apple

- Realplayer 8 <http://www.real.com/> by RealNetworks

- Soja, <http://www.helio.org> a Java based SMIL player by Helio

- S2M2 <http://smil.nist.gov/player> , a Java Applet-based SMIL Player by NIST

- Schmunzel <http://www.salzburgresearch.at/suntrec/schmunzel/> , a Java player by SunTREC Salzburg.

- X-SMILES <http://www.xsmiles.org/> a Java based open browser by TML laboratory

3.366 AUTHORING TOOLS:

- Ezer <http://www.smilmedia.com> by SMIL Media

- Fluition <http://www.confluenttechnologies.com> by Confluent Technologies

- Grins <http://www.oratrix.com/GRiNS/index.html> by Oratrix

- GoLive6 <http://www.adobe.com/products/golive/overview.html> by Adobe

- HomeSite <http://www.allaire.com/products/homesite/index.cfm> by Allaire

- MAGpie <http://ncam.wgbh.org/webaccess/magpie> , a captioning tool by WGBH

- Hi-Caption <http://www.hisoftware.com/hmcc/acc4mcc.html>, a captioning tool by Hisoftware
- MovieBoard <http://www.simple.co.jp/products/10MovieBorad.htm>, for e-learning (Japanese only)

- MMS Simulators <http://lists.w3.org/Archives/Public/www-mobile/2002Aug/0007.html> list

- Perly SMIL <http://www.webiphany.com/perlysmil/> , a SMIL 1.0 Perl module

- RealSlideshow Basic <http://forms.real.com/rnforms/products/tools/slideshowbasic/index.html?key=868E21032182964> by RealNetworks

- SMIL Composer SuperToolz <http://autodownload.sausage.com> by HotSausage

- Smibase <http://smibase.com/>, a server-installed software suite

- SMIL Editor V2.0 <http://www.docomo-sys.co.jp/prod/soft/smil2.html>, by DoCoMo.

- SMILGen <http://www.smilgen.org> by RealNetworks, a SMIL (and XML) authoring tool designed to ease the process of XML.

- SMIL Scenario Creator <http://w3-mcgav.lab.kdd.co.jp/sc/indexe.html> by KDDI

- TAG Editor 2.0 - G2 release <http://tag.digital-ren.com> by Digital Renaissance ???

- Tagfree 2000 SMIL Editor <http://www.tagfree.com/english/product/product02.asp?menu=2>

- TransTool <http://www.psych.uiuc.edu/~kmiller/dvguide/analysis_tools.htm> - open source transcription tool

- VeonStudio <http://www.veon.com/> by Veon

- Validator: SMIL 1.0, SMIL 2.0, SMIL 2.0 Basic and XHTML+SMIL <http://www.cwi.nl/~media/symm/validator/> by CWI.
- SMG <http://www.smilmedia.com.> for a PDA, a BREW, a Phone and a PC by Smilmedia

- The IBM Toolkit for MPEG-4 <http://www.alphaworks.ibm.com/tech/tk4mpeg4> creates MPEG-4 binary from content created in XMT-O (based on the SMIL 2.0 syntax and semantics).

SMIL's critics say it overlaps and even conflicts with existing standards, including HTML 4.0, Dynamic HTML, Cascading Style Sheets (CSS) and in particular, the Document Object Model (DOM).

Finally better impact, improved accessibility, and easier maintenance add up to better user and better developer experience. Unlike my wife's professor, I don't believe that SMIL will replace HTML. But it should at least displace HTML from the task of serving up unsynchronized video in pop-up windows. Whether SMIL makes its mark via standalone players such as RealPlayer and QuickTime or via browsers like IE 5.5, folks will smile when they surf to your SMIL enhanced site.

3.37 DOCUMENT OBJECT MODEL (DOM):

3.371 Introduction:

The Document Object Model is a platform- and language-neutral interface that will allow programs and scripts to dynamically access and update the content, structure and style of documents. The document can be further processed and the results of that processing can be incorporated back into the presented page. This is an overview of DOM-related materials here at W3C and around the web.

3.372 Why the Document Object Model?

"Dynamic HTML" is a term used by some vendors to describe the combination of HTML, style sheets and scripts that allows documents to be animated. The W3C has received several submissions from members companies on the way in which the object model of HTML documents should be exposed to scripts. These submissions do not propose any new HTML tags or style sheet technology. The W3C DOM Working Group is working hard to make sure interoperable and scripting-language neutral solutions are agreed upon.

W3C's Document Object Model (DOM) is a standard Application Programming Interface (API) to the structure of documents; it aims to make it easy for programmers to access components and to delete, add, or edit their content, attributes and style. In essence, the DOM makes it possible for programmers to write applications that work properly on all browsers and servers and on all platforms. While programmers may need to use different programming languages, they do not need to change their programming model. W3C's DOM thus offers programmers a platform and language neutral program interface, which will make

programming reliably across platforms with languages such as Java and ECMA Script a reality.

It is the most widely implemented XML parser interface today. DOM reproduces an XML document's data hierarchy in a programming language's native object format, giving programmers an easy and familiar way of working with the data in the document. Developers can iterate through the document's data elements and even change the document's content programmatically.

However, the DOM recommendation does not cover searching or file input/output (loading and saving XML documents). The DOM API loads the entire XML document into memory, favoring repetitive operations performed on short documents. For lengthy documents, the SAX (Simple API for XML) API is a better choice.

DOM Level 2 Brings Platform-Neutral Dynamic Content to the Web Created and developed by the W3C Document Object Model (DOM) Working Group, this specification extends the platform- and language-neutral interface to access and update dynamically a document's content, structure, and style first described by the DOM Level 1 Recommendation. The DOM Level 2 provides a standard set of objects for representing Extensible Markup Language (XML) documents and data, including namespace support, a style sheet platform which adds support for CSS 1 and 2, a standard model of how these objects may be combined, and a standard interface for accessing and manipulating them.

Leading the Web to its full potential, the World Wide Web Consortium (W3C) released the Document Object Model Level 2 specification as a W3C Recommendation. The specification reflects cross-industry agreement on a standard API (Applications Programming Interface) for manipulating documents and data through a programming language (such as Java or ECMA Script). A W3C Recommendation indicates that a specification is stable, contributes to Web interoperability, and has been reviewed by the W3C Membership, who favor its adoption by the industry.

This W3C recommendation extends the basic data representation capability of the DOM API to other programming concepts such as custom namespaces, style sheets, events, iterators, filters, and range functions. This gives developers a standardized way of expressing functions that they previously had to create on their own. For example, if you want your document to call a particular function whenever it encounters a specific event in the data, DOM Level 2 provides a simple and standard way of doing so.

The DOM Level 2 Recommendation builds on the solid work done in DOM Level 1, and gives Web authors the power to move to XML for dynamic content," says Lauren Wood of SoftQuad Software Inc., and Chair of the W3C DOM Working Group. "The DOM also provides developers with the interoperability and integration ability they need. There are now several implementations of the DOM, in different programming languages, which provide the basis of powerful systems meeting the business needs of several large organizations.

Most commercial XML processing software include support only for namespaces and style sheets, so using a product that supports DOM Level 2 will give developers a greater range of flexibility. However, because the other types of triggers are not widely supported, one could run into compatibility problems that negatively affect the portability of his project.

DOM Level 2 Delivers Interoperable Software for XML Documents with Namespace Support DOM Level 1 was designed for HTML 4.0 and XML 1.0. With DOM Level 2, authors can take further advantage of the extensibility of XML. Simply put, anywhere you use XML, you can now use the DOM to manipulate it.

The standard DOM interface makes it possible to write software (similar to plug-ins) for processing customized tag-sets in a language- and platform-independent way. A standard API makes it easier to develop modules that can be re-used in different applications. DOM Level 2 provides support for XML namespaces, extending and improving the XML platform.

As more sites move to XML for content delivery, DOM Level 2 emerges as a critical tool for developing dynamic Web content.

DOM Level 2 Extends the Dynamic, Device Independent Web The DOM defines a standard API that allows authors to write programs that work without changes across tools and browsers from different vendors. But beyond this, it provides a uniform way to produce programs that work across a variety of different devices, so all may benefit from dynamically generated content.

3.373 DOM Architecture:

The DOM Architecture is divided into various modules. Each module addresses a particular domain. Domains covered by the current DOM API are XML, HTML, Cascading Style Sheets (CSS), and tree events. Future domains can be the rendered content (that is, the content displayed on the screen which might differ from the input document), user agent function, etc.

3.3731 DOM Core:

The DOM Core defines a tree-like representation of the document, also referred as the DOM tree, enabling the user to traverse the hierarchy of elements accordingly.
Refer also to the DOM Range and Traversal modules to manipulate the tree elements/structure defined in the DOM Core.

3.3732 DOM XML:

The XML DOM extends the Core platform for specific XML 1.0 needs, such as processing instructions, CDATA, and entities.

3.3733 DOM HTML:

The HTML DOM defines a set of convenient easy to use ways to manipulate HTML documents. The initial HTML DOM only describes methods, such as how to access an

identifier by name, or a particular link. The HTML DOM is sometimes referred to as *DOM Level 0* but has been imported into *DOM Level 1*.

3.3734 DOM Events:

This part defines XML-tree manipulation oriented events with tree mutation and user-oriented events such as mouse, keyboard, and HTML-specific events.

3.3735 DOM Cascading Style Sheets:

The DOM CSS defines a set of convenient, easy to use ways to manipulate CSS style sheets or the formatting of documents.

3.3736 DOM Load and Save:

Loading an XML document into a DOM tree or saving a DOM tree into an XML document is a fundamental need for the DOM user. This module includes a variety of options controlling load and save operations.

3.3737 DOM Validation:

This module defines a set of methods to modify the DOM tree and still make it valid.

3.3738 DOM Xpath:

The DOM XPath defines a set of convenient, easy to use functions to query a DOM tree using an XPath 1.0 expression, such as evaluate.

3.374 DOM Requirements:

The DOM requirements <DOMTR> contains all requirements for each Level, and is regularly updated to reflect the requirements of the latest Level.

3.3741 DOM Level 0:

Functionalities equivalent to the ones exposed in Netscape Navigator 3.0 and Microsoft Internet Explorer 3.0 are informally referred to as "Level 0". There is no W3C specification for this Level.

3.3742 DOM Level 1:

DOM Level 1 <DOMTR> was completed in October 1998 and provides support for XML 1.0 </TR/REC-xml> and HTML 4.0 </TR/html4>.

3.3743 DOM Level 2:

DOM Level 2 <DOMTR> was completed in November 2000 extending Level 1 with support for XML 1.0 with namespaces </TR/REC-xml-names/>, adding supports for Cascading Style Sheets </TR/CSS2> (CSS), events such as user interface events and tree manipulation events, and enhancing tree manipulation methods (tree ranges and traversal mechanisms). Level 2 HTML is a W3C Recommendation since January 2003.

3.3744 DOM Level 3:

DOM Level 3 <DOMTR> is currently under development. Level 3 will extend Level 2 by finishing support for XML 1.0 with namespaces aligning the DOM Core with the XML Infoset </TR/xml-infoset/>, adding support for XML Base </TR/xmlbase/>, and extending the user interface events (keyboard). Level 3 will also add support for validation, the ability to load and save a document, explore further mixed markup vocabularies and their implications on the DOM API ("Embedded DOM"), and will support XPath </TR/xpath>.

Note: The DOM Working Group <Group> has released a public Working Draft of the Views and Formatting </TR/2000/WD-DOM-Level-3-Views-20001115/> model. The DOM Working Group <Group> has released a W3C Note on the Abstract Schemas </TR/2002/NOTE-DOM-Level-3-AS-20020725/> model. This document is no longer a work item of the Working Group.

3.3745 Other DOMs:

The DOM Working Group is not the only Working Group within the W3C to produce APIs and extensions to the DOM architecture. Other DOM modules include:

- DOM for MathML 2.0 </TR/MathML2/>: generic API for MathML 2.0 documents.

- DOM for SMIL Animation </TR/smil-animation/>: generic API for SMIL animation.

- DOM for SVG 1.0 </TR/SVG/>: generic API for SVG 1.0 documents.

3.375 DOM Test Suites:

The W3C DOM Activity is developing the DOM Conformance Test Suites <Test> in coordination with NIST <http://www.nist.gov> (National Institute of Standards and Technology) and the public community, with help from a few W3C Members. A first version <Test> was released in February 2002.

3.376 DOM Working Group Licensing mode:

The W3C DOM Working Group is a royalty-free Working Group, as defined in the Current Patent Practice Note <http://www.w3.org/TR/2002/NOTE-patent-practice-20020124>. A list of disclosures from the Working Group participants <http://www.w3.org/2002/08/02-DOM-Disclosures.html> is available.

At present the W3C is not aware of any patents that are essential to implement the DOM specifications. Therefore, it is the W3C's opinion that the DOM specifications can be implemented on a royalty-free basis.

3.38 SCALABLE VECTOR GRAPHICS (SVG):

3.381 Introduction:

Over the past few years, there has been a demonstrated interest in vector graphics and animation on the Web. In 1999, the World Wide Web Consortium (W3C) started developing an open format called Scalable Vector Graphics (SVG). The first SVG 1.0 Specification was published in September 2001.

Scalable Vector Graphics (SVG) offers Web developers a method to create and animate images through an XML programming language. Consequently, rather than being removed from their code as is often the case with proprietary technology, developers can gain finer degrees of control over the appearance of Web pages. Animation techniques can range from a simple linear movement to 3D double helix morphing effects. Web developers, once they are more aware of the possibilities, can find unprecedented levels of control.

It is open source and offers seamless database integration, server-scripting compatibility, and efficient accessibility/localization workflow. SVG is a language for describing two-dimensional graphics in XML. SVG allows for three types of graphic objects: vector graphic shapes (e.g., paths consisting of straight lines and curves), images and text. Graphical objects can be grouped, styled, transformed and composited into previously rendered objects. Text can be in any XML namespace suitable to the applications, which enhances searchability and accessibility of the SVG graphics. The feature set includes nested transformations, clipping paths, alpha masks, filter effects, template objects and extensibility.

SVG drawings can be dynamic and interactive. The Document Object Model (DOM) for SVG, which includes the full XML DOM, allows for straightforward and efficient vector graphics animation via scripting. A rich set of event handlers such as onmouseover and onclick can be assigned to any SVG graphical object. Because of its compatibility and leveraging of other Web standards, features like scripting can be done on SVG elements and other XML elements from different namespaces simultaneously within the same Web page.

3.382 Features:

There are many advantages of using SVG as the following short feature list demonstrates:

- Compatibility with other mediums such as wireless devices

- Scalable Server Solutions

- Small file sizes for faster Web page downloads

- Unlimited color and font choices
- Zoomable graphics and images
- Scripting control for custom interactive events and animation
- Clean, crisp, high-resolution printing from Web browsers
- Bitmap-style filter effects for high-impact graphics
- Text-based format easily integrates with other Web technologies
- Built in International Language Support
- Reduced Maintenance Costs
- Easily Updated
- Rich Multimedia Capabilities

Besides these important features, one of the best features of SVG is its Human readability. From the earliest days developers have been examining existing HTML and JavaScript to learn how to write new content, to make improvements, and to develop a better Web. SVG is intentionally following the same mode. Flash designer Joshua Davies of Praystation.com fame releases some of his source code and developers all over the world love it (the results are delivered as a non-human, readable binary format). Intentionally, every SVG graphic has its source visible and is human readable. Power to the developers through "View Source." Not only does human readability have benefits for content creators and developers; it has enormous benefits to accessibility. Jacek asks: "Is staring at the SVG code going to make the message that the image is carrying easier to understand?" The answer is a

resounding yes! The text within an SVG image is just that text. Anyone can read the text, where it is located, and how it will be rendered by simply looking at the source. An image described as four wheels, a body, some doors, and seats is probably a vehicle of some kind. Someone that has trouble seeing the image, for whatever reason, has a chance at finding the meaning by reading the source. Another example is user style sheets. Suppose users cannot view a graphic because they are color-blind. Since they can read the source, they can easily write a user style sheet to override the colors in the graphic.

Similar Web based technology, Flash, the Web standard for animation and vector graphics is also available but SVG is all time better than Flash. Flash and SVG are often compared because the two have similar features. The reality is that SVG has some distinct advantages over its main competitor Flash. Perhaps chief among them is the compliance with other standards. SVG can utilize CSS and the DOM, where as Flash relies on proprietary technology that is not open source, at least not in the sense that we can right click on the page and see what is happening behind the scenes. SVG by contrast is open source and developers can readily learn from other developer's efforts in this area. While SVG has not yet reached the popularity level of Flash, times are changing quickly bringing with it a sense of enthusiasm for SVG.

While discussing about the populrity and acceptability of SVG, it is important to know that Mozilla plans to fully support SVG, Microsoft has similar plans, and Adobe GoLive 5 also supports SVG. Additionally, SVG editors are now surfacing on the Web. Programs such as Jasc's WebDraw that allows for the creation of SVG in a visual format are excellent additions to the SVG paradigm. The SVG 1.0 Recommendation has significant support from industrial giants some of them are Adobe, Apple, Canon, Corel, Hewlett-Packard, Macromedia, Microsoft, Kodak, Sun and many others, who contributed to the specification. There are a large number of different SVG viewer implementations. The most popular is the Adobe browser plug-in/ActiveX control. Adobe has released its viewer on all the versions of Windows that Microsoft supports (98, 2000, Millennium, and XP), Macintosh

OS versions 8.5 to X, Linux, and Solaris. It runs in Internet Explorer, Netscape, Mozilla, Opera, and many other browsers.

3.383 Drawbacks:

There are however some drawbacks in this technology, one of the major drawbacks at the moment are:

- No browser fully supports SVG currently. As a consequence, SVG has to be displayed through the use of a plug-in such as the Adobe SVG plug-in <http://www.adobe.com/svg/>. While it is a good plug-in it does not currently support all the SVG specifications, it is a heavy download, and perhaps the biggest barrier is that it is CPU intensive. Still, despite these drawbacks it does allow for cross-browser implementation of SVG and the use of the plug-in is likely to increase dramatically in the years to come.

- The Adobe SVG Viewer relies on JavaScript for most of its dynamic and interactive features, so it will not work in Internet Explorer 5 for Macintosh. Critical bugs were found running the Adobe SVG Viewer with Netscape on both Macintosh and PC.

- The other drawback is that there is a distinct lack of online material that directly relates to understanding how SVG can be employed in developing Web sites.

Extended example titled "Half Steppin'," expands on the ideas presented in the simple project. Because SVG is open source, one can examine the code for Half Steppin' to learn how the techniques have been applied.

After analysing various features and drawbacs it can be concluded, as the future of SVG not only seems bright, it seems certain to play a major role on the Web in the years to come.

3.391 XFORM:

3.3911 Introduction:

The World Wide Web Consortium (W3C) announced the release of the first Public Working Draft of the XForms Data Model in 1999. The XForms Data Model Working Draft, along with the XForms Requirements document, provide the first cross-industry efforts in seven years to produce the next generation of Web-based forms. W3C is building a better Web Form when HTML Forms were introduced to the Web in 1993, they provided a means to gather information and perform transactions. The structure of forms served the needs of many users at that time, as well as the devices used to access the Web.

Seven years later, the Web is a space where hundreds of millions of users expect to use many different devices to perform increasingly complex transactions, many of which exceed the limitations of the original forms technology. The W3C HTML Working Group has a charter to develop a form architecture that provides a better match to workflow and database applications, to the proliferation of new Web-enabled devices, and to the XML-driven Web.

The XForms Subgroup has accepted the challenge and produced a form architecture that separates data modeling, logic, and presentation. The XForms Data Model has emerged as the first in a series of XForms specifications.

The current design of Web forms doesn't separate the *purpose* from the *presentation* of a form. XForms, in contrast, are comprised of separate sections that describe what the form does, and how the form looks. This allows for flexible presentation options, including classic XHTML forms, to be attached to an XML form definition. The following illustrates how a single device-independent XML form definition, called the **XForms Model**, has the apability to work with a variety of standard or proprietary user interfaces:

- The **XForms User Interface** provides a standard set of visual controls that are targeted toward replacing today's XHTML form controls. These form controls are directly usable

inside XHTML and other XML documents, like SVG. Other groups, such as the Voice Browser Working Group, may also independently develop user interface components for XForms.

- An important concept in XForms is that forms collect data, which is expressed as XML **instance data**. Among other duties, the XForms Model describes the structure of the instance data. This is important, since like XML, forms represent a structured interchange of data. Workflow, auto-fill, and pre-fill form applications are supported through the use of instance data.

Finally, there needs to be a channel for instance data to flow to and from the XForms Processor. For this, the **XForms Submit Protocol** defines how XForms send and receive data, including the ability to suspend and resume the completion of a form.

3.3912 Key Goals of Xforms:

- Support for handheld, television, and desktop browsers, plus printers and scanners
- Richer user interface to meet the needs of business, consumer and device control applications
- Decoupled data, logic and presentation
- Improved internationalization
- Support for structured form data
- Advanced forms logic
- Multiple forms per page, and pages per form

- Suspend and Resume support
- Seamless integration with other XML tag sets

3.3913 XForms Data Model Separates Purpose from Presentation:

XForms aims to ease the transition of the Web from HTML to XML. As XHTML 1.0 allows HTML content authors to make a smooth entry into the XML world, XForms allow Web application authors to combine the modularity of XML with the simplicity of HTML to gain key advantages in the areas of device independence, accessibility, business-to-business and consumer e-commerce, and embedded devices.

The XForms Data Model deliberately separates the purpose of a form from its presentation. This allows the application author to rigorously define the form data, independent of how end-users interact with the application. The separation facilitates the development of Web applications with user interaction components, and provides advantages to Web application developers.

3.3914 XForms Deliver Structured Data, Device Independence:

In the XForms suite of specifications, the rules for describing, validating, and submitting application data are expressed in XML, as well as the submitted data. By providing the rules and data in XML, XForms lays the foundation for combinations with other XML applications, supporting the extensible Web.

Separating purpose and presentation also makes device independence easier to achieve by allowing Web application authors to write the data model once for all devices. Because the data model is not tied to presentation, developers may customize the presentation in a way that best suits each device's user interface. Support for device independence paves the way for a Web that is accessible to all users.

XForms Implementations, Drafts in Progress:

The XForms subgroup is producing early implementations of XForms, to determine requirements and test ideas for the specification. Examples are available from the XForms page. Other members of the subgroup have committed to implementing XForms in their products.

The XForms Data Model is the first in a series of XForms specifications. Other XForms work focuses on the logic layer - identifying relationships and dependencies between data model fields - and on the presentation aspects.

3.392 Xpath:

3.3921 Introduction:

XPath is the result of an effort to provide a common syntax and semantics for functionality shared between XSL Transformations [XSLT] and XPointer [XPointer]. The primary purpose of XPath is to address parts of an XML document. In support of this primary purpose, it also provides basic facilities for manipulation of strings, numbers and booleans. XPath uses a compact, non-XML syntax to facilitate use of XPath within URIs and XML attribute values. XPath operates on the abstract, logical structure of an XML document, rather than its surface syntax. XPath gets its name from its use of a path notation as in URLs for navigating through the hierarchical structure of an XML document.

In addition to its use for addressing, XPath is also designed so that it has a natural subset that can be used for matching (testing whether or not a node matches a pattern); XPath models an XML document as a tree of nodes. There are different types of nodes, including element nodes, attribute nodes and text nodes. XPath defines a way to compute a string-value for each type of node. Some types of nodes also have names. XPath fully supports XML Namespaces. Thus, the name of a node is modeled as a pair consisting of a local part and a possibly null namespace URI; this is called an expanded-name.

The primary syntactic construct in XPath is the expression. An expression is evaluated to yield an object, which has one of the following four basic types:

- node-set (an unordered collection of nodes without duplicates)
- boolean (true or false)
- number (a floating-point number)
- string (a sequence of UCS characters)

Expression evaluation occurs with respect to a context. XSLT and XPointer specify how the context is determined for XPath expressions used in XSLT and XPointer respectively. The context consists of:

- a node (the context node)
- a pair of non-zero positive integers (the context position and the context size)
- a set of variable bindings
- a function library
- the set of namespace declarations in scope for the expression

The context position is always less than or equal to the context size.
The variable bindings consist of a mapping from variable names to variable values. The value of a variable is an object, which can be of any of the types that are possible for the value of an expression, and may also be of additional types not specified here.

The function library consists of a mapping from function names to functions. Each function takes zero or more arguments and returns a single result. This document defines a

core function library that all XPath implementations must support. For a function in the core function library, arguments and result are of the four basic types. Both XSLT and XPointer extend XPath by defining additional functions; some of these functions operate on the four basic types; others operate on additional data types defined by XSLT and XPointer.

The namespace declarations consist of a mapping from prefixes to namespace URIs. The variable bindings, function library and namespace declarations used to evaluate a subexpression are always the same as those used to evaluate the containing expression. The context node, context position, and context size used to evaluate a subexpression are sometimes different from those used to evaluate the containing expression. Several kinds of expressions change the context node; only predicates change the context position and context size. When the evaluation of a kind of expression is described, it will always be explicitly stated if the context node, context position, and context size change for the evaluation of subexpressions; if nothing is said about the context node, context position, and context size, they remain unchanged for the evaluation of subexpressions of that kind of expression.

XPath expressions often occur in XML attributes. The grammar specified in this section applies to the attribute value after XML 1.0 normalization. So, for example, if the grammar uses the character <, this must not appear in the XML source as < but must be quoted according to XML 1.0 rules by, for example, entering it as <. Within expressions, literal strings are delimited by single or double quotation marks, which are also used to delimit XML attributes. To avoid a quotation mark in an expression being interpreted by the XML processor as terminating the attribute value the quotation mark can be entered as a character reference (" or '). Alternatively, the expression can use single quotation marks if the XML attribute is delimited with double quotation marks or vice-versa.

One important kind of expression is a location path. A location path selects a set of nodes relative to the context node. The result of evaluating an expression that is a location path is the node-set containing the nodes selected by the location path. Location paths can recursively contain expressions that are used to filter sets of nodes.

3.3922 Boolean Functions:

The boolean function converts its argument to a boolean as follows:

- a number is true if and only if it is neither positive or negative zero nor NaN
- a node-set is true if and only if it is non-empty
- a string is true if and only if its length is non-zero
- an object of a type other than the four basic types is converted to a boolean in a way that is dependent on that type

Function: boolean not(boolean)

The not function returns true if its argument is false, and false otherwise.

Function: boolean true()

The true function returns true.

Function: boolean false()

The false function returns false.

Function: boolean lang(string)

The lang function returns true or false depending on whether the language of the context node as specified by xml:lang attributes is the same as or is a sublanguage of the language specified by the argument string. The language of the context node is determined by the value of the xml:lang attribute on the context node, or, if the context node has no xml:lang attribute, by the value of the xml:lang attribute on the nearest ancestor of the context node that has an xml:lang attribute. If there is no such attribute, then lang returns false. If there is such an attribute, then lang returns true if the attribute value is equal to the argument ignoring case, or if there is some suffix starting with - such that the attribute value is equal to the argument ignoring that suffix of the attribute value and ignoring case.

3.3923 Number Functions:

The number function converts its argument to a number as follows:

- a string that consists of optional whitespace followed by an optional minus sign followed by a Number followed by whitespace is converted to the IEEE 754 number that is nearest (according to the IEEE 754 round-to-nearest rule) to the mathematical value represented by the string; any other string is converted to NaN

- boolean true is converted to 1; boolean false is converted to 0

- a node-set is first converted to a string as if by a call to the string function and then converted in the same way as a string argument

- an object of a type other than the four basic types is converted to a number in a way that is dependent on that type

If the argument is omitted, it defaults to a node-set with the context node as its only member.

3.3924 Data Model:

XPath operates on an XML document as a tree. This section describes how XPath models an XML document as a tree. This model is conceptual only and does not mandate any particular implementation. The relationship of this model to the XML Information Set [XML Infoset] is described in [B XML Information Set Mapping].

XML documents operated on by XPath must conform to the XML Namespaces Recommendation [XML Names].

The tree contains nodes. There are seven types of node:

- root nodes
- element nodes
- text nodes

- attribute nodes
- namespace nodes
- processing instruction nodes
- comment nodes

For every type of node, there is a way of determining a string-value for a node of that type. For some types of node, the string-value is part of the node; for other types of node, the string-value is computed from the string-value of descendant nodes.

3.393 CASCADING STYLE SHEETS:

3.3931 Introduction:

Cascading Style Sheets (CSS) is a simple mechanism for adding style (e.g. fonts, colors, spacing) to Web documents.

3.3932 CSS Browsers:

The easiest way to start experimenting with style sheets is to download one of the browsers that support CSS. Not all of the browsers below implement the full specification, but releases are coming out fast so this should soon change. Various sites describe bugs and work-arounds.

- The KDE <http://www.kde.org> project released KDE 3.1, which includes the Konqueror Web browser and file manager. It has improved support for CSS 2.1, including fixed table layout and positioning.
- Opera <http://www.opera.com/> released version 7 <http://www.opera.com/products/desktop/index.dml?platform=windows> of its browser, with some new CSS-based goodies: small-screen mode, alternative styles, etc. (Opera 6 runs on multiple platforms, version 7 is so far Windows-only, shareware)
- Apple <http://www.apple.com/> released a beta of the Safari Web browser <http://www.apple.com/safari/>. It uses KHTML <http://www.konqueror.org/konq-

browser.html> (from the KDE <http://www.kde.org/> project) as rendering engine. (free, Mac OS X)

- Mozilla <http://www.mozilla.org/> released version 1.1 <http://www.mozilla.org/releases/> and Netscape <http://home.netscape.com/> version 7.0 <http://channels.netscape.com/ns/browsers/download.jsp>, based on Mozilla 1.0.1. Both have excellent CSS support. (Mozilla is Open Source, Netscape is binary-only but free, both run on many platforms)
- The Chimera project <http://chimera.mozdev.org/> released version 0.4 <ftp://ftp.mozilla.org/pub/chimera/releases/chimera-0.4.dmg.gz>. Chimera is a browser for Mac OS X, based on Mozilla's Gecko layout engine. (Mac, Open Source)
- The X-Smiles team <http://www.x-smiles.org/> has released version 0.5 ("Oulu") of the X-Smiles XML browser, which supports <http://www.xsmiles.org/xsmiles_features.html>, among other things, XHTML, SMIL, Xforms and the CSS Mobile Profile. (Java, Open Source)
- NetClue <http://www.netcluesoft.com/> released Clue Browser v4.1.1. It supports HTML, XML/XHTML, namespaces, CSS (level 1 and part of level 2), DOM, Javascript, etc. (Java)
- Microsoft released Internet Explorer for the Mac 5.1 <http://www.microsoft.com/mac/download/ie/ie51.asp>, with bug fixes and improved performance. Supports full CSS1 and partial CSS2. (Mac IE 5 was the first browser to reach better than 99% support for CSS1 <http://www.webreview.com/style/css1/leaderboard.shtml>, in March 2000.) (free; Mac OS 8, 9 & X)
- OmniWeb 4 <http://www.omnigroup.com/applications/omniweb/> is a Web browser for the Mac (OS X) and has a built-in source editor (with HTTP PUT support). (Shareware)
- Galeon 1.0 <http://galeon.sourceforge.net/> is a Web browser for Gnome <http://www.gnome.org>. It uses the Gecko <http://www.mozilla.org/newlayout/> rendering engine from Mozilla <http://mozilla.org> internally. (Open Source, Unix)

- Adobe <http://www.adobe.com/> produces an SVG plugin <http://www.adobe.com/svg/viewer/install/> for browsers under Mac and Windows and for Mozilla 0.9.1 under Linux & Solaris <http://www.adobe.com/svg/viewer/install/old.html>. Supports SVG with CSS styling. (free)
- K-Meleon <http://kmeleon.sourceforge.net/> version 0.6 has been released, a lightweight browser based on the Gecko <http://www.mozilla.org/newlayout/> rendering engine of Mozilla (Windows, Open Source)
- Espial's <http://www.espial.com> Escape 4.7 browser <http://www.espial.com/main/page?view=p-escp_main> implements CSS support for HTML, XML and XHTML. Written in Java for the embedded software market.
- iCab <http://www.icab.de>, a browser for the Mac, is starting to support CSS. The preview release of version 2.5 reportedly supports most of CSS1. (Free)
- Openwave's <http://www.openwave.com/> mobile browser <http://www.openwave.com/products/browser.html> implements XHTML and CSS and is expected to ship in cell phones 2nd half of 2001. Also see data sheet [PDF] <http://www.openwave.com/resources/docs/Openwave_MobileBrowser.pdf>.
- Nokia <http://www.nokia.com/> will start selling mobile phones that support XHTML and CSS during 2001. See demo [Flash] <http://www.nokia.com/xhtmldemo/>, press release <http://press.nokia.com/PR/200103/813189_5.html> and white paper [PDF] <http://www.nokia.com/press/background/pdf/mar011.pdf>.
- The Arachne WWW browser <http://www.arachne.cz/> for DOS and Linux supports CSS1 since version 1.70 (free for non-commercial use).
- CSIRO <http://www.cmis.csiro.au/> released the CSIRO SVG Toolkit <http://sis.cmis.csiro.au/svg/>, with a viewer for SVG + CSS and other utilities. (Java, Open Source)
- IONIC <http://www.ionicsoft.com/index.html> offers the Ionic SVG toolkit <http://www.ionicsoft.com/ionic/svg/index.html>, with a viewer for SVG + CSS and other tools. (Java)

- The Koala team wrote Jackaroo <http://koala.ilog.fr/jackaroo/>, an SVG + CSS viewer. (Jackaroo has now merged with Batik and is no longer supported.) (Java, Open Source)
- Microsoft <http://www.microsoft.com> shipped Internet Explorer 5 for the Macintosh. <http://www.microsoft.com/mac/ie> It apparently supports *full* CSS1, the first browser to do so.
- Closure <http://www.uni-karlsruhe.de/~unk6/closure/> is a Web browser written in Common Lisp; supports CSS1.
- Hewlett Packard <http://www.hp.com/> released their "embedded microbrowser" ChaiFarer <http://www.chai.hp.com/chai_farer.html>, supporting CSS1. CSS2 will come later.
- ICE Soft <http://www.icesoft.com/> released v.5 of their two embeddable browsers <http://www.icesoft.com/ICEBrowser/index.html>: the "base" one is a viewer for HTML/XML+CSS2, the "pro" one adds networking and more. Both in Java. Does MathML, too.
- Microsoft has released Internet Explorer 5.0 for Windows, Solaris and HP-UX <http://www.microsoft.com/windows/ie/default.htm>
- Silicon Graphics has an embeddable CSS-enhanced web browser that is used in a number of applications and their desktop
- Arena <http://www.yggdrasil.com/Products/Arena/>, previously W3C's testbed browser, is now being developed by Yggdrasil <http://www.yggdrasil.com>. It has a partial implementation of CSS1.
- Emacs-w3 <http://www.cs.indiana.edu/elisp/w3/docs.html>, a.k.a. Gnuscape Navigator, supports some CSS1.

These sources document the level of support in various browsers:

- Johannes Koch has a nice page with work-arounds for various browser bugs <http://w3development.de/css/hide_css_from_browsers/summary/>.

- The SVG working group has a detailed list of the features <../../Graphics/SVG/Test/BE-ImpStatus-20011026.html> (including CSS support) of various SVG implementations.
- RichInStyle.com <http://www.richinstyle.com/> has lists of bugs <http://www.richinstyle.com/bugs/> for various browsers. (Careful: as of Oct 2001, there are still several bugs in the list of bugs itself.)
- Western Civilisation <http://www.westciv.com/> compares CSS1 support <http://www.westciv.com/style_master/academy/browser_support/index.html> in Netscape, Internet Explorer and Opera.
- Do you have fear of style sheets <http://www.alistapart.com/stories/fear/>? Jeffrey Zeldman has the cure.
- The CSS Pointers Group <http://css.nu> documents CSS bugs <http://www.css.nu/pointers/bugs.html> in major browsers.
- ProjectCool <http://www.projectcool.com> documents CSS properties <http://www.projectcool.com/developer/cssref/ref.html> and tells you what works in which browser <http://www.projectcool.com/developer/reference/css_style.html>.
- WebReview <http://www.webreview.com/>'s The Browser Compatibility Chart <http://www.webreview.com/style/css1/charts/mastergrid.shtml> is a thorough review of how the implementations match the specification.
- Braden N. McDaniel has documented MS IE3 for Windows95/NT <http://www.shadow.net/~braden/nostyle/ie3.html>

3.3933 CSS Specifications:

Cascading Style Sheets, level 1 (CSS1) </TR/REC-CSS1> became a W3C Recommendation <../../TR/> in December 1996. It describes the CSS language as well as a simple visual formatting model. CSS2 </TR/REC-CSS2>, which became a W3C Recommendation in May 1998, builds on CSS1 and adds support for media-specific style sheets (e.g. printers and aural devices), downloadable fonts, element positioning and tables. The CSS Mobile Profile <../../TR/css-mobile> specification became a W3C Candidate Recommendation <../../TR/> in Oct 2001.

CSS3 is currently under development. Progress may be checked at <current-work> as new drafts are published.

Translations into some languages are available from the CSS1 translations page <../css1-updates/translations> and the CSS2 translations page <../css2-updates/translations>. *Errata* are maintained separately for CSS1 <../css1-updates/REC-CSS1-19990111-errata> and CSS2 <../css2-updates/REC-CSS2-19980512-errata>.

3.394 EXTENSIBLE STYLE LANGUAGE (XSL):

3.3941 Introduction:

The World Wide Web Consortium (W3C) has published the first working draft of its Extensible Style Language (XSL) 1.0 in 1998. XML brought new features which were unavailable in HTML without necessarily making HTML obsolete. The idea is that web authors should have various tools to hand for each job, one simple to learn and straightforward to use, the other perhaps less simple but with greater functionality and extensibility so as to be readily customizable. HTML and CSS are the basic tools, with XML and XSL intended as their industrial-strength cousins. The W3C goes on the say that CSS, already a mature standard, and XSL are to be based on a single formatting model. The two will share the same underlying concepts and use the same terminology as far as is possible. So what has XSL got that CSS hasn't? It can handle tree transformations as well as document transformations, it permits XML documents to be displayed in different ways in response to different user queries, and it can support many languages, including historical texts like ancient Greek and Aztec. Ian Jacobs of W3C explains"It's also designed to be used for print a little bit more than CSS". In addition, where CSS works with HTML or XML, XSL is optimized for use with XML. It does seem odd, though, that the W3C is out pushing a new stylesheet language when the newly formed Web Standards Group (CI No 3,471) is complaining that browser vendors aren't adhering closely enough to existing web standards.Microsoft has posted Extensible Style Sheet Language (XSL) tools on its Web site to enable users to experiment with the technology as it works its way through the standards process.

XSL enables structured Extensible Markup Language (XML) data to be formatted for the Web. The XSL proposal was authored by Microsoft, ArborText, and Inso. Whitehill Technologies Inc., a leading provider of Internet infrastructure software, today announced the second major release of their leading edge, standard-setting XSL software, Whitehill (xsl) Composer, including the addition of a robust XSL editor. Whitehill (xsl) Composer is a desk top development application that revolutionizes the ability for an organization to capitalize on XML based business communications. Using existing XML data, Whitehill (xsl) Composer automatically creates XSL/CSS (cascading style sheets), which can be used to render web versions of electronic bills, invoices, statements and reports as well as for XML/EDI data exchange using XSLT (extensible style sheet transforms). Until now, the creation of XSL to render XML electronically has been a completely manual process. Whitehill (xsl) Composer replaces the need to hand code XSL.

With the addition of a robust XSL Editor, Whitehill (xsl) Composer provides the user with the ability to import and edit existing XSL, as well as providing an Absolute Positioning feature, which allows users to position and align fragments in relation to the page layout or to other fragments. This functionality significantly increases the speed at which users create the presentation of their XML data while leveraging any pre-existing investment in XSL.

The World Wide Web Consortium (W3C) has issued the Extensible Stylesheet Language (XSL) 1.0 as a W3C Recommendation, representing cross-industry agreement on an XML-based language that specifies how XML documents may be formatted. It works in concert with XSL Transformations (XSLT), an XML language that performs transformations of structured documents.

3.3942 XSL 1.0 brings Structured Styling to XML Documents:

For document-driven industries, the Extensible Markup Language (XML) has held great promise, but also presented some limitations. While XML has proven an effective format for structured data, it had yet to provide the advanced levels of formatting and structural transformation common to proprietary publishing tools.

XSLT 1.0, the XML language which performs transformations on XML data and documents, has been a W3C Recommendation since November 1999, and already enjoys significant usage in both developer communities and in commercial products. XSL 1.0 builds on XSLT 1.0, and provides users with the ability to describe how XML data and documents are to be formatted. XSL 1.0 does this by defining "formatting objects," such as footnotes, headers, columns, and other features common to paged media.

Designers would use XSL 1.0 stylesheets to indicate rendering preferences for a type of XML document, including how it is styled, laid out, and paginated onto a presentation medium such as a browser window, a pamphlet, or a book. An XSL engine would take the XML document and the XSL stylesheet, and would produce a rendering of the document. XSLT 1.0 makes it possible to significantly change the original structure of an XML document (automatic generation of tables of contents, cross-references, indexes, etc.), while XSL 1.0 makes complex document formatting possible through the use of formatting objects and properties.

3.3943 XSL 1.0 enriches XML Documents and Data with Professional Printing Capabilities:

As XSL 1.0 is focused on the formatting of paged media, it makes it possible for professional printing capabilities and functions to perform with XML documents today. XSL 1.0 and XSLT make it possible for the needs of Web and print-based media formatting to be met. Now, one can have documents and data stored in XML, specify how to format and render them, and produce versions for both Web rendering and for print media.

3.3944 XSL 1.0 Complements CSS Technologies:

The Cascading Style Sheet language (CSS), both levels 1 and 2 has long been recognized as the style language of choice for HTML and XHTML documents. CSS may still be used for XML formatting, and in cases where structural transformations are not needed, suit the needs of Web designers.

The W3C CSS and XSL Working Groups have cooperated to ensure that their results are complementary. Using CSS properties and the CSS formatting model, the XSL Working Group has ensured complete compatibility and interoperability between the two families for styling.

XSL Benefits from Industry Support and User Testing Key industry leaders and XML experts participated in the creation of both the transformation and formatting components of XSL, including (in alphabetical order) Adobe, Antenna House, Arbortext, Bitstream, Enigma, IBM, James Clark, Microsoft, Oracle, RivCom, SoftQuad, Software AG, Sun Microsystems, University of Edinburgh, and Xerox. Implementation commitments are significant, and are included in the testimonials for XSL 1.0.

3.395 URI, URL and URN

3.3951 Introduction:

The Web is an information space. Human beings have a lot of mental machinery for manipulating, imagining, and finding their way in spaces. URIs are the points in that space. Unlike web data formats, where HTML is an important one, but not the only one, and web protocols, where HTTP has a similar status, there is only one Web naming/addressing technology: URIs.

3.3952 Uniform Resource Identifiers:

URIs, URLs are short strings that identify resources in the web: documents, images, downloadable files, services, electronic mailboxes, and other resources. They make resources available under a variety of naming schemes and access methods such as HTTP, FTP, and Internet mail addressable in the same simple way. They reduce the tedium of "log in to this server, then issue this magic command ..." down to a single click.
It is an extensible technology: there are a number of existing addressing, and more may be incorporated over time.

3.3953 URL (Uniform Resource Locator):

An informal term (no longer used in technical specifications) associated with popular URI schemes: http, ftp, mailto, etc.

3.3954 URN (Uniform Resource Name):

An URI that has an institutional commitment to persistence, availability, etc. Note that this sort of URI may also be a URL. See, for example PURLs. A particular scheme, URN, specified by RFC2141 and related documents, intended to serve as persistent, location-independent, resource identifiers. Engelbart also identified the need for a library system, including catalog numbers for documents. Catalog numbers are a long-standing tradition in publishing and library science. The technology behind Uniform Resource Identifiers is suitable for use as catalog numbers, given sufficient socioeconomic infrastructure: rights management, payment assurance, privacy, digital signatures, etc.

3.396 ACTIVE SERVER PAGE (ASP):

3.3961 Introduction:

To search and display dynamic and live information, such as that found in locally produced relational databases, requires the addition of other languages. In such case the systems should use Active Server Pages (ASP). Before discussing it is necessory to know about dynamic web page and the static web page.

3.3962 Static Web Pages:

It is easy to create Web pages to display static data from a database table. Most systems, such as Microsoft Access, include a "Save As HTML" feature to do just that. However, pages created in this way are a "snapshot in time"; they do not change as the database is changed. It is essential to re-create the pages each time changes are made to the data. Publishing and maintaining a large number of static Web pages is a maintenance is worse, users cannot search the database or choose particular items of data; they can only see what it is previously saved in HTML format.

3.3963 Dynamic Web Pages:

Dynamic Web pages allow the user to connect to up-to-the-minute data, search it, and display it in different ways. It contains following features:

3.39631 Client-side processing:

There are two ways of incorporating processing into a Web page: client-side processing and server-side processing. In client-side processing, the server sends a Web page containing both code and data back to the client (user's computer). The browser gets the whole thing and proceeds to call up programs on the user's computer to process the data. The user's computer does all the work; the server just supplies the code and data.

Client-side processing has traditionally been characterized by programs, or scripts, written in languages such as Java, JavaScript, or Visual Basic Scripting Edition (VBScript). Microsoft's VBScript is a subset of the Visual Basic language which is used in both client-side and server-side processing.

Access 2000 has a new object and method for creating client-side access to its databases. It is called a "data access page", and automates all the programming to convert Access tables into dynamic Web pages. The disadvantage is that it imbeds an Access program into the Web page and the user must have a significant amount of software on his or her computer to make use of it (Viescas, 1999).

3.39632 Server-side processing:

The other way is to use server-side processing. Here, the server calls up programs residing on the server. The programs process the data and send the processed data back to the user's browser as HTML for display. This puts the processing load on the server, and it must have necessary programs installed. The server does all the work.

3.39633 Common Gateway Interface (CGI):

An early method of server-side processing, still widely used, is CGI, or Common Gateway Interface. It is a standard against which to write programs for Web access. Programs written according the CGI standard take requests from a Web page, retrieve and process data from a database, and send the processed data back to the user's browser (Khurana and Khurana, 1966).

CGI programs may be written in almost any language that can be compiled: Visual Basic (not VBScript), C, Perl, and so on. A complete CGI setup usually includes a requesting HTML page, a CGI program, and a response HTML page to display the data. The CGI program searches and manipulates the database then creates lines of HTML code in which data are embedded. The program finally writes a file of HTML code to be sent to the browser. The file is read by the browser as a regular Web page. A good Web-based tutorial on CGI is by Selena Sol (1998).

Since CGI programs reside on the server, not on the user's computer, you may have to jump through hoops to get the network administrator to install a CGI program. For security reasons, some administrators will not permit CGI programs to be used at all.

3.3964 Active Server Pages:

A simpler method is to use Active Server Pages (ASP). ASP is Microsoft's method of providing server-side processing for use by Web browsers. ASP imbeds scripting statements directly into the Web page, rather than by using a separate program. ASP differs from CGI in the following ways:

- ASP is proprietary to Microsoft and CGI is not. ASP requires the use of Microsoft server software, and works best in a complete Microsoft environment; CGI can be used on most any computer platform, and is widely used on UNIX servers.
- ASP usually uses VBScript whereas CGI programs may be written in any language that can be compiled. Although VBScript may be used in client-side processing, it is always server-side in an ASP context.

- ASP imbeds the script within the Web page, while CGI programs stand alone, separate from the Web page that calls them.

- Unless you are a programmer, it is much easier to use ASP than CGI.

For most applications, CGI is slower than ASP.

ASP uses ActiveX Data Objects (ADO), which is a Microsoft-developed technology for working with databases on the Internet. They are pre-packaged programming libraries that are built into the Web server and called upon to do some work when it needs to be done. ADO relieves the programmer from writing code for every little thing that he/she wants to do. Not much useful work would get done if a painter must first grab a camel, pluck its hairs, make a brush, fashion a bucket out of tin, and construct a ladder each time he/she wanted to paint a wall. Instead, he/she grabs the tools and materials off the shelf and concentrates on the painting project. ASP programmers can reach into a set of pre-built tools (ADO) and concentrate on the larger picture of working with data (Kauffman, 1999).

ASP technology has many uses, but this article will concentrate on using it to search and retrieve data from a Microsoft Access database. Other texts (Francis et al., 1998) and Web tutorials (Caroll, n.d.) describe non-database ASP applications. Typically, the browser requests an AS page from the server, either by URL or through the use of a Web form[1]. The server sees the .asp extension, and sends the AS page to the ASP engine on the server. The engine reads the file, which usually contains a mixture of server side scripting and HTML. As the ASP engine reads the file, whenever it comes to server-side script, it executes it. Whenever it comes to HTML, it outputs it back to the server (Kauffman, 1999, p. 224).

Access has a built-in utility for exporting an Access table, query, form, or report as an Active Server Page. It is limited in its functionality, although it may be a good place to start. This article will discuss creating ASP from scratch. The benefit of ASP is that the full power of a relational database is available to the user, with the Web providing the interface. The

user needs nothing more than a browser. Typically, the user points his or her browser to a menu on your Web site which links to various AS pages that work with the database.

3.397 LIBWWW- THE W3C PROTOCOL LIBRARY:

3.3971 Introduction:

Libwww is a highly modular, general-purpose client side Web API written in C for Unix and Windows (Win32). It is well suited for both small and large applications, like browser/editors, robots, batch tools, etc. Pluggable modules provided with libwww include complete HTTP/1.1 (with caching, pipelining, PUT, POST, Digest Authentication, deflate, etc), MySQL logging, FTP, HTML/4, XML, RDF, WebDAV, and much more. The purpose of libwww is to serve as a testbed for protocol experiments.

Rough consensus and running code is the main idea behind libwww. As for all W3C OpenSource code, the purpose of libwww is to provide an environment for experimenting with extensions and new features. The focus of libwww is performance, modularity, and extensibility. It contains highly efficient code for HTTP and URIs but also for many other parts of the Web, primarily for client side applications like robots, browsers, GUI apps, and automated tools.

Designed and implemented libwww from version 2.17 up to version 5.2.8 Tim Berners-Lee and Jean-Francois Groff. Who brought up with the initial design and implementation of libwww. Libwww has been part of the World Wide Web almost from the beginning. Tim Berners-Lee designed and implemented the first version back in November 1992 as part of demonstrating the potential of the Web. Many people have picked up libwww and used it in a variety of contexts. Applications such as Lou Montulli's Lynx character based client, Mosaic Web browser by Marc Andreesen and Eric Bena, and the CERN server by Ari Luotonen were all using later versions of libwww. Later on, applications like the Arena browser by Dave Raggett and Håkon W. Lie have been added to the list.

Libwww was free from the very start and was released on a regular basis to the Web Community. When CERN stopped being the center of the Web in late 1994, libwww moved

from CERN which continued its development. In May 1998, the code base was made even more available in that people now can check it out directly from CVS codebase. Today, libwww is freely available under W3C Copyright for use by anyone and has a growing OpenSource community helping maintaining it. Along with the core library comes set of sample applications that demonstrate how to use libwww but at the same time can perform useful tasks in their own right.

3.398 MULTI MODAL INTRACTION:

3.3981 Introduction:

Web pages you can speak to, a new class of mobile devices that support multiple modes of interaction. And which contains following features:

- Adapting the Web to allow multiple modes of interaction:
  - GUI, Speech, Vision, Pen, Gestures, Haptic interfaces, ...

- Augmenting human to computer and human to human interaction
  - Communication services involving multiple devices and multiple people

- Anywhere, Any device, Any time
  - Services that adapt to the device, user preferences and environmental conditions

- Accessible to all

The Multimodal Interaction Activity is extending the Web user interface to allow multiple modes of interaction, offering users the choice of using their voice, or an input device such as a key pad, keyboard, mouse, stylus or other input device. For output, users will be able to listen to spoken prompts and audio, and to view information on graphical displays. The Working Group is developing markup specifications for synchronization across multiple modalities and devices with a wide range of capabilities.

3.3982 Current Situation:

The Multimodal Interaction Activity was chartered in February 2002 as a royalty free group under W3C's Current Patent Practice (CPP) Note. The following organizations are currently participating in the Working Group:

Alcatel, Apple, AT&T, Avaya, Canon, Cisco, Comverse, Corel, EDS, Ericsson, France Telecom, Hewlett-Packard, IBM, Intel, IWA/HWG, Kirusa, Loquendo, Microsoft, Mitsubishi Electric, Motorola, NEC, Nokia, Nortel Networks, Nuance Communications, OnMobile Systems, Opera Software, Openstream, Oracle, Panasonic, PipeBeach, ScanSoft, Siemens, SnowShore Networks, SpeechWorks International, Sun Microsystems, T-Online International, Toyohashi University of Technology, V-Enable, VoiceGenie, and Voxeo

All participating organizations are required to make a patent disclosure statement as set out in the CPP Note. A separate page is being maintained for patent disclosures for the Multimodal Interaction Activity. The Working Group is obliged by its charter to produce a specification which relies only on intellectual property available on a royalty-free basis.

3.3983 Current Drafts:

- Multimodal Interaction Framework (6 May 2003). This introduces a general framework for multimodal interaction, and the kinds of markup languages being considered.

- Multimodal Interaction Use Cases (4 December 2002). This describes several use cases that are helping us to better understand the requirements for multimodal interaction.

- Multimodal Interaction Requirements (8 January 2003). Describes fundamental requirements for the specifications under development in the W3C Multimodal Interaction Activity.

- EMMA Requirements (13 January 2003). Describes requirements for a data format (EMMA) that acts as an exchange mechanism between input processors and interaction management components in a multimodal application.
- Ink Requirements (22 January 2003). Describes requirements for a data format for representing ink entered with an electronic pen or stylus in a multimodal system. The markup will allow for the input and processing of handwriting, gestures, sketches, music and other notational languages in web-based multimodal applications.

3.3984 Background:

3.39841 Current Devices:

Desktop systems have proven to be highly effective for accessing the World Wide Web. The high resolution displays, pointing devices and full size keyboards make it easy to interact efficiently with large amounts of information. When you are on the move, you need a small lightweight device that fits easily into your pocket or purse. Cell phones are extremely popular, but their small size limits the amount of information they can display, as well as the number and kinds of keys they can feature.

Mobile profiles have emerged for a number of W3C specifications: XHTML, CSS, SMIL and SVG. Mobile access to the Web is now becoming a reality. The small keypads make it difficult to enter search strings or Web addresses, especially for ideographic languages with many thousands of characters. Recent years have also seen a tremendous growth of interest in using speech as a means to interact with Web-based services over the telephone. W3C responded to this by establishing the Voice Browser Activity which is developing requirements and specifications for the W3C Speech Interface Framework.

Spoken interfaces based upon VoiceXML prompt users with pre-recorded or synthetic speech and understand simple words or phrases. As the technology improves we can look forward to richer natural language conversations. There is now an emerging interest in combining speech interaction with other modes of interaction. Multimodal interaction will

enable the user to speak, write and type, as well as hear and see using a more natural user interface than today's single mode browsers.

3.39842 Multimodal Access:

The different modalities may be supported on a single device or on separate devices working in tandem, for example, you could be talking into your cellphone and seeing the results on a PDA. Voice may also be offered as an adjunct to browsers with high resolution graphical displays, providing an accessible alternative to using the keyboard or screen. This can be especially important in automobiles or other situations where hands and eyes free operation is essential. Voice interaction can escape the physical limitations on keypads and displays as mobile devices become ever smaller. It is much easier to say a few words than it is to thumb them in on a keypad where multiple key presses may be needed for each character. Complementing speech, ink entered with a stylus or imaging device can be used for handwriting, gestures, drawings, and specific notations for mathematics, music, chemistry and other fields. Ink is expected to be popular for instant messaging.

Mobile devices working in isolation generally lack the power to recognize more than a few hundred spoken commands. The storage limitations restrict the use of prerecorded speech prompts. Small speech synthesizers are possible, but tend to produce robotic sounding speech that many users find tiring to listen to. A solution is to process speech recognition and synthesis remotely on more powerful platforms. A similar case holds for complex voice dialogs with rich natural language understanding. Simple dialogs could be handled locally, but for richer interaction, it will be necessary to couple the device with a remote dialog engine.

Multimodal applications should be able to adapt to changing device capabilities, user preferences and environmental conditions. For instance, users should be able to disable speech input and output when this would be distracting to nearby people. It should be easy for developers to tailor applications to dynamically adapt to such changes, making best use of the available modes of interaction at any given time. In addition, developers should be able to

create applications involving multiple devices and multiple users, augmenting human to computer and human to human interaction.

The Multimodal Interaction Working Group is chartered to work on developing standards for synchronization across modes and devices, building on top of W3C's existing specifications, for instance, combining XHTML, SMIL and XForms with markup for speech synthesis and speech recognition. Alternatively it can provide mechanisms for loosely coupling visual interaction with voice dialogs represented in VoiceXML. Additional work will focus on a means to provide the ink component of Web-based, multimodal applications.

3.39843 Extensible Multimodal Annotation Markup Language (EMMA):

- First Working Draft - expected late July 2003

- Last Call Working Draft - TBD

EMMA is being developed as a data format for the interface between input processors and interaction management systems. It will define the means for recognizers to annotate application specific data with information such as confidence scores, time stamps, input mode (e.g. key strokes, speech or pen), alternative recognition hypotheses, and partial recognition results etc. EMMA is a target data format for the semantic interpretation specification being developed in the Voice Browser Activity, and which describes annotations to speech grammars for extracting application specific data as a result of speech recognition. EMMA supercedes earlier work on the natural language semantics markup language in the Voice Browser Activity.

3.39844 Pen input:

- First Working Draft - expected July 2003

- Last Call Working Draft - expected last quarter 2003

This work item sets out to define an XML data format for ink entered with an electronic pen or stylus as part of a multimodal system. This will enable the capture and server-side processing of handwriting, gestures, drawings, and specific notations for mathematics, music, chemistry and other fields. The starting point for this work is a specification contributed by IBM, Intel, the International Unipen Foundation and Motorola. This will be reviewed with respect to the requirements for ink in the context of multimodal applications.

3.3985 The working Process:

3.39851 Interpreting and Representing the User's Input:

Simple natural language processing is needed to transform natural language input (whether from speech, pen or keystrokes) so that it can be used to fill in forms, follow hypertext links, shift the focus of interaction and so on. Simple natural language processing is also useful for dynamically generating tailored visual and aural responses. One approach to implementing this is to combine XHTML </MarkUp/> with markup for prompts, grammars and the means to bind results to actions. XHTML </MarkUp/> defines various kinds of events, for example, when the document is loaded or unloaded, when a form field gets or loses the input focus, and when a field's value is changed. These events can in principle be used to trigger aural prompts, and to activate recognition grammars. This would allow a welcome message to start playing when the page is loaded. When you set the focus to a given field, a prompt could be played to encourage the user to respond via speech rather than via keystrokes. Page grammars could be used for navigation and for switching tasks. Grammars are not restricted to speech, and ink grammars could be used for recognizing gestures and characters entered with a stylus.

3.39852 Collaborative Processing between Local and Network Devices:

Cell phones are expected to include support for an increasing number of specs: XHTML, CSS, SMIL, SVG, SyncML, ECMAScript, Java, JPEG, MP3, and more. At the same time, the costs must be held down to a low enough level to make the phones affordable to people from all walks of life. This together with battery considerations constrains the

available memory and processor speed. This motivates looking at an architecture where the speech and natural language processing is done in the network. In addition, this will encourage innovation, for example on richer natural language capabilities, without the need for everyone to upgrade their phones to enjoy the benefits.

The IETF Speech Services Control (SpeechSC) working group is developing protocols to support distributed speech recognition, speech synthesis and speaker verification services, and expects to take advantage of W3C's work on the speech recognition grammar specification (SRGS), the speech synthesis markup language (SSML), semantic interpretetation (SI) and extensible multimodal markup annotations (EMMA).

Another idea is to couple a local graphical user interface with a remote voice dialog engine, perhaps based upon VoiceXML. Here the idea is to allow events to be passed between the device and the remote dialog engine. To the application developer, these events would look just the same whether they originated locally or remotely. The IETF SIP working group is developing a means for pass such events over SIP (session initiation protocol). In this model, events can carry data, and can thus be used to initiate a range of actions, for instance, changing the focus of interaction, setting the value of a form field, loading a new page, or altering the current page via the DOM.

SIP can also be used to synchronize several devices, for instance to update the display on a PDA, automotive or desktop system in concert with the much smaller display on a cellphone. When it comes to setting up a session that potentially involves multiple devices and servers, SIP looks like it will provide an effective solution together with server-side scripts.

3.3991 PLATFORM FOR INTERNET CONTENT SELECTION (PICS):

3.39911 Introduction:

The PICS specification enables labels (metadata) to be associated with Internet content. It was originally designed to help parents and teachers control what children access

on the Internet, but it also facilitates other uses for labels, including code signing and privacy. The PICS platform is one on which other rating services and filtering software have been built.

In August of 1995, leading members of the Internet community came together to begin the development of technical specifications that would enable users to 1) easily find appropriate content and 2) avoid content that they consider inappropriate or unwanted, either for themselves or their children. These specifications were designed to ease the creation of, and access to, labeling schemes (and associated content selection and filtering mechanisms), allowing various people or organizations to label Web content in ways that best suit their different viewpoints. The PICS specifications were not intended to be limited to applications regarding potentially offensive content. Rather, it was hoped that PICS would be used for many purposes, such as third-party ratings on the timeliness and technical accuracy of a site's content.

Final technical specifications were completed in early 1996. Since then PICS has been incorporated into a number of products <http://www.w3.org/PICS/>, a variety of PICS-based rating services <http://www.w3.org/PICS/> have been (and continue to be) developed for the web, and a number of stand-alone filtering tools <http://www.microsys.com/pics/software.htm> are PICS-compatible.

Many who were involved in the creation of PICS recognized that the World Wide Web provides access to an extraordinary range of content, some of which some people consider either inappropriate, unwanted, or harmful for some users, especially children. The global nature of the Web, and the fact that it serves numerous communities with a great diversity of values, suggested that national, or even international laws restricting certain kinds of speech on the Web would neither be effective nor necessarily desirable for the Web. Instead, PICS was developed to accommodate a wide range of communities online.

The original PICS proposers based their work on a general set of principles, detailed below. In the time since PICS and other content selections tools have been deployed on the web, much has been learned about the use of PICS-based techniques. This note builds on those Principles a set of functional guidelines for implementing PICS-based components of the Web infrastructure, PICS rating services, and PICS-based content selection tools to assure that they are designed in a manner that comports with the original PICS Principles and the free flow of information on the Web.

3.39912 PICS Statement of Principles:

PICS is a cross-industry working group whose goal is to facilitate the development of technologies to give users of interactive media, such as the Internet, control over the kinds of material to which they and their children have access. PICS members believe that individuals, groups and businesses should have easy access to the widest possible range of content selection products, and a diversity of voluntary rating systems.

In order to advance its goals, PICS will devise a set of standards that facilitate the following:

3.399121 Self-rating:

Enable content providers to voluntarily label the content they create and distribute.

3.399122 Third-party rating:

Enable multiple, independent labeling services to associate additional labels with content created and distributed by others. Services may devise their own labeling systems, and the same content may receive different labels from different services.

3.399123 Ease-of-use:

Enable parents and teachers to use ratings and labels from a diversity of sources to control the information that children under their supervision receive.

PICS members believe that an open labeling platform which incorporates these features provides the best way to preserve and enhance the vibrancy and diversity of the Internet. Easy access to technology which enables first- and third-party rating of content will give users maximum control over the content they receive without requiring new restrictions on content providers.

Membership in PICS includes a broad cross-section of companies from the computer, communications, and content industries, as well as trade associations and public interest groups. PICS member will deploy products and services based on these standards.

3.39913 Guidelines for the Usage of PICS:

In addition to the principles above, we recommend that systems and services based on PICS ought to be implemented with the following guidelines in mind. These guidelines promote the principles of diversity, disclosure, control, and transparency.

- **Using PICS Rating Systems and Services:**

  -The Web, through PICS implementations, ought to support access to a variety of labeling systems that reflect the diversity of moral and cultural values held by those that use the Net.

  -No single rating system and service ~~can~~ perfectly meet the needs of all the communities on the web.

  -The ability of multiple organizations to use PICS to create lists of suggested content is an encouraged means of using PICS. These lists may be distributed through label bureaus and be used for searching, or as "white" lists of materials that should be permitted even if they would otherwise be blocked.

  -Filtration and labeling schemes should be designed such that the combined effect does not lead to a chilling of expression or the creation of significant barriers to diverse opinion and content. Small and non-commercial sites should continue to be a part of the Web available to all users.

- **Creating Labeled Content:** The creation of content that is labeled should be done in a way so as to maximize the transparency and integrity of the Web.

  -PICS-based systems should facilitate disclosure of the criteria used to rate content.

  Content rating should be as simple as possible for authors and content providers who wish to label content.

  -The decision to self-label should be at the discretion of content creators and publishers.

  -If a content creator is concerned about the accuracy of a third party rating, she should be able to investigate how her materials are rated and have some means of requesting a change in the ratings where they do not match the stated criteria of the rating service.

- **Using Labeled Content:** Users should have the ability to understand and control the choices made in the selection of content in an easy and transparent manner.

  -Users of PICS-based content selection systems should have easy access to information about the filtering criteria, the values or principles underlying them, and to the configuration of the content selection systems. This can be accomplished by providing the following information in the product documentation or at the Service URL:

  -a clear statement of the methodology used to create the labels;

  -a contact (both physical and virtual) for questions or concerns.

When access to a particular URL is blocked through an implementation of PICS, error conditions or other user interface functions ought to specifically indicate that the URL is not accessible because of blocking by a content selection tool. Relevant information could include:

-the rating system whose value is out of range (if more than one is being used) and which variable and value led to the blocking of a URL.

-some indication of where the blocking occurred.(i.e. is it part of the browser and under local control, or is it a proxy and if so who owns and/or operates the proxy.)

-It should be as easy as possible for an authorized user to install and modify filters.

-In particular, we recommend that filtering software have the ability to import filtering preferences that are specified using the PICSRules language.

3.39914 Resources for Labeling Service Developers:

To start a new labeling service, it is needed to take the following steps:

-Decide who will assign labels.

- Web site operators who self-label *and/or*
- A panel of raters that you recruit *and/or*
- A computer program that analyzes the contents of materials and assigns labels

-Decide the labeling vocabulary and criteria

-Express the labeling vocabulary and criteria according to the format specified in the technical specification <http://www.w3.org/TR/REC-PICS-services>. You can create this file from scratch, or you can fill out web forms at the PICS Application Incubator <http://www.si.umich.edu/~presnick/PICS-incubator/> and the file will be created for you.

-Create the labels

-Arrange for distribution of your labels

- Give your labels to someone else who is running a PICS label bureau *and/or*
- Run your own PICS label bureau *and/or*

- Convince web site operators to distribute the labels for their own pages, either by putting them into HTML META tags or sending them along with web pages.

The PICS Application Incubator <http://www.si.umich.edu/~presnick/PICS-incubator/> project at the University of Michigan School of Information will provide a limited amount of free technical consulting to organizations that are considering establishing new labeling services.

3.39915 PICS Capabilities:

3.399151 PICS can be used for more than just content filtering:

While the motivation for PICS was concern over children accessing inappropriate materials, it is a general "meta-data" system, meaning that labels can provide any kind of descriptive information about Internet materials. For example, a labeling vocabulary could indicate the literary quality of an item rather than its appropriateness for children. Most immediately, PICS labels could help in finding particularly desirable materials (see, for example, NetShepherd's label-informed Alta Vista search <http://www.netshepherd.com/>), and this is the main motivation for the ongoing work on a next generation label format that can include arbitrary text strings. More generally, the W3C <http://www.w3.org> is working to extend Web meta-data capabilities generally and is applying them specifically in the following projects:

Digital Signature Project<http://www.w3.org/pub/WWW/Security/DSig/Overview.html>
coupling the ability to make assertions with a cryptographic signature block that ensures integrity and authenticity.

Intellectual Property Rights Management <http://www.w3.org/pub/WWW/IPR/>
using a meta-data system to label Web resources with respect to their authors, owners, and rights management information.

Privacy (P3) <http://www.w3.org/pub/WWW/Privacy/Overview.html>

using a meta-data system to allow sites to make assertions about their privacy practices, and for users to express their preferences for the type of interaction they want to have with those sites.

Regardless of content control, meta-data systems such as PICS are going to be an important part of the Web, because they enable more sophisticated commerce (build and manage trust relationships), communication, indexing, and searching services.

The promise of digital commerce is that it will allow you to use the Internet to purchase the services of the best organic gardening advisors or mad cow disease specialists, whether they live in Santa Clara or Timbuktu. To do this, you need to do more than verify that the person at the other end of the wire is who he says he is. You need to assess competence, reliability, judgment. In other words, you need a system of branding, but applied much more widely for highly specialized and hard-to-evaluate services and products. You need value-added services that will not only lead you to the right product or service but also rate its quality or otherwise vouch for it.

3.399152 PICS enables censorship:

This seemingly straightforward question, upon closer inspection, turns out to be many different questions when asked by different people. Many people are concerned about governments assuming one or more of the roles described in the answer to the previous question. Others are concerned about employers setting filtering rules, abuse of power by independent labelers, or a chilling effect on speech even if speech is not banned outright. People also employ different definitions of censorship. The most expansive definition is, "any action by one person that makes otherwise available information unavailable to another person." Under this expansive definition, even a parent setting filtering rules for a child would count as censorship. PICS documents have adopted the more restrictive definition of censorship as actions that limit what an individual can distribute, and use the term "access controls" for restrictions on what individuals can receive. But the distinction blurs if a central

authority restricts access for a set of people. Finally, people have different definitions of "enable." Some would say that PICS enables any application that uses PICS-compatible components, while we reserve the term "enables" for applications that can easily be implemented with PICS-compatible components but could not be easily implemented otherwise.

Given the variety of implicit questions, it doesn't make sense to provide a blanket answer to the question of whether PICS enables censorship. This FAQ answers many of the specific questions that people often mean when they ask the more general question. For example, we ask questions about whether PICS makes it easier or harder for governments to impose labeling and filtering requirements. If you believe there's another specific question that should be addressed, please send it to pics-ask@w3.org <mailto:pics-ask@w3.org>, for possible inclusion in a later version.

3.399153 PICS makes it easier or harder for governments to do so:

A government could try to assume any or all of the six roles described above, although some controls might be harder than others to enforce. As described below, governments could assume some of these roles even without PICS, while other roles would be harder to assume if PICS had not been introduced. It's important to note that W3C does not endorse any particular government policy. The purpose of this FAQ is to explain the range of potential policies and to explore some of the impacts of those policies on both the climate of intellectual freedom and the technical infrastructure of the World Wide Web. Potential government policies:

Set labeling vocabulary and criteria. A government could impose a labeling vocabulary and require all publishers (in the government's jurisdiction) to label their own materials according to that vocabulary. Alternatively, a government might try to achieve the same effect by encouraging an industry self-policing organization to choose a vocabulary and require subscribers to label their own materials. Civil liberties advocates in Australia are especially concerned about this (see The Net Labeling Delusion

<http://www.pobox.com/~rene/liberty/label.html>). PICS makes it somewhat easier for a government to impose a self-labeling requirement: without PICS, a government would have to specify a technical format for the labels, in addition to specifying the vocabulary and criteria, and there might not be any filtering software available that could easily process such labels.

Assign labels. A government could assign labels to materials that are illegal or harmful. This option is most likely to be combined with government requirements that such materials be filtered (see #5 below) but it need not be; a government could merely provide such labels as an advisory service to consumers, who would be free to set their own rules, or ignore the labels entirely. If a government merely wants to label, and not impose any filtering criteria, then PICS again provides some assistance because it enables a separation of labeling from filtering. On the other hand, a government that wishes to require filtering of items it labels as illegal gets little benefit from PICS as compared to prior technologies, as discussed below in the question about national firewalls.

Distribute labels. A government could operate or finance operation of a Web server to distribute labels (a PICS label bureau); the labels themselves might be provided by authors or independent third parties. Taken on its own, this would actually contribute to freedom of expression, since it would make it easier for independent organizations to express their opinions (in the form of labels) and make those opinions heard. Consumers would be free to ignore any labels they disagreed with. Again, since PICS separates labeling from filtering, it enables a government to assist in label distribution without necessarily imposing filters. If combined with mandatory filtering, however, a government-operated or financed label bureau could contribute to restrictions on intellectual freedom.

Write filtering software. It's unlikely that a government would write filtering software rather than buying it; the supplier of filtering software probably has little impact on intellectual freedom.

Set filtering criteria. A government could try to impose filtering criteria in several ways, including government-operated proxy servers (a national intranet), mandatory filtering by service providers or public institutions (e.g., schools and libraries), or liability for possession of materials that have been labeled a particular way. In some ways, by enabling independent entities to take on all the other roles, PICS highlights this as the primary political battleground. Each national and local jurisdiction will rely on its political and legal process to answer difficult policy questions: Should there be any government-imposed controls on what can be received in private or public spaces? If so, what should those controls be? Most kinds of mandatory filters could be implemented without PICS. A government could express its required filtering criteria in the form of a PICSRule that everyone would be required to install and run, but without PICSRules a government could express its requirements in less technical form. One potential policy, however, mandatory filtering based on labels provided by non-government sources, would have been difficult to impose without PICS.

Install/run filters. A Government could require that filtering software be made available to consumers, without mandating any filtering rules. For example, a government could require that all Internet Service Providers make filtering software available to its customers, or that all PC browsers or operating systems include such software. Absent PICS, governments could have imposed such requirements anyway, since proprietary products such as SurfWatch and NetNanny are available.

3.399154 PICS encourages individual controls rather than government controls:

For example, a national proxy-server/firewall combination that blocks access to a government-provided list of prohibited sites does not depend on interoperability of labels and filters provided by different organizations. While such a setup could use PICS-compatible technology, a proprietary technology provided by a single vendor would be just as effective. Other controls, based on individual or local choices, benefit more from mixing and matching filtering software and labels that come from different sources, which PICS enables. Thus, there should be some substitution of individual or local controls for centralized controls, although it is not obvious how strong this substitution effect will be. In Europe initial calls

for centralized controls gave way to government reports calling for greater reliance on individual recipient controls; the end results of these political processes, however, are yet to be determined.

3.399155 Labeling

It matters whether labels are applied to IP addresses or to URLs:

An IP address identifies the location of a computer on the Internet. A URL identifies the location of a document. To simplify a little, a URL has the form http://<domain-name>/<filename>. A web browser first resolves (translates) the domain-name into an IP address. It then contacts the computer at that address and asks it to send the particular filename. Thus, a label that applies to an IP address is a very broad label: it applies to every document that can be retrieved from that machine. Labeling of URLs permits more flexibility: different documents or directories of documents can be given different labels. This difference of granularity will, naturally, have an impact on filtering. Filters based on IP addresses will be cruder: if some but not all of the documents available at a particular IP address are undesirable, the filter will have to either block all or none of those documents. PICS, by contrast, permits labeling of individual URLs, and hence permits finer grain filters as well.

3.3991551 Self-labeling

3.39915511 PICS makes author self-labeling more effective:

Without a common format for labels, authors could not label themselves in a way that filtering programs could make use of. PICS provides that format.

3.39915512 PICS makes a government requirement of self-labeling more practical to implement:

It enables such a requirement to have more impact. A government requirement of self-labeling would have little impact if the labels were not usable by filtering programs. PICS provides the common format so that filtering software from one source can use labels provided by other sources (authors in this case).

3.39915513 Self-labeling depend on universal agreement on a labeling vocabulary and criteria for assigning labels to materials:

Although universal agreement is not necessary, there does need to be some harmonization of vocabulary and labeling criteria, so that labels provided by different authors can be meaningfully compared.

3.39915514 PICS makes it easier for governments to cooperate in imposing self-labeling requirements:

PICS provides a language-independent format for expressing labels. If governments agreed on a common set of criteria for assigning labels, the criteria could be expressed in multiple languages, yet still be used to generate labels that can be compared to each other.

3.39915515 It is effective for (some) authors to label their own materials as inappropriate for minors. What about labeling appropriate materials?

Both kinds of labeling could be effective, but only if a high percentage of the materials of a particular type are labeled. If the inappropriate materials are labeled, then a filter can block access to the labeled items. If the appropriate materials are labeled, then a filter can block access to all the unlabeled items.

3.3991552 Third-party labeling

3.39915521 Can an organization I dislike label my web site without my approval?

Yes. Anyone can create a PICS label that describes any URL, and then distribute that label to anyone who wants to use that label. This is analogous to someone publishing a review of your web site in a newspaper or magazine.

3.39915522 Isn't there a danger of abuse if a third-party labeler gets too powerful?

If a lot of people use a particular organization's labels for filtering, that organization will indeed wield a lot of power. Such an organization could, for example, arbitrarily assign negative labels to materials from its commercial or political competitors. The most effective way to combat this danger is to carefully monitor the practices of labeling services, and to

ensure diversity in the marketplace for such services, so that consumers can stop using services that abuse their power.

3.3991553 Other Social Concerns About Labeling

3.39915531 Why did PICS use the term "label", with all of its negative associations?

PICS documents use the term "label" broadly to refer to any machine-readable information that describes other information. Even information that merely classifies materials by topic or author (traditional card catalog information) would qualify as labels if expressed in a machine-readable format. The PICS developers recognized that the term "label" has a narrower meaning, with negative connotations, for librarians and some other audiences, but it was the most generic term the PICS creators could find without reverting to technical jargon like "metadata."

In media with centralized distribution channels, such as movies, labeling and filtering are not easily separated. For example, unrated movies are simply not shown in many theaters in the USA. In addition to its technical contribution, PICS makes an intellectual contribution by more clearly separating the ideas of labeling and filtering. Many of the negative connotations associated with "labeling" really should be associated with centralized filtering instead. There are, however, some subtle questions about the impact of labeling itself, as articulated in the next two questions.

3.39915532 Does the availability of labels impoverish political discussions about which materials should be filtered?

Matt Blaze (personal communication) describes this concern with an analogy to discussions at local school board meeting about books to be read in a high school English class. Ideally, the discussion about a particular book should focus on the contents of the book, and not on the contents of a review of the book, or, worse yet, a label that says the book contains undesirable words.

There will always be a tradeoff, however, between speed of decision-making and the ability to take into account subtleties and context. When a large number of decisions need to be made in a short time, some will have to be made based on less than full information. The challenge for society, then, will be to choose carefully which decisions merit full discussion, in which case labels should be irrelevant, and which decisions can be left to the imperfect summary information that a label can provide. The following excerpt from Filtering the Internet <http://www.sciam.com/0397issue/0397resnick.html> summarizes this concern and the need for eternal vigilance:

> "Another concern is that even without central censorship, any widely adopted vocabulary will encourage people to make lazy decisions that do not reflect their values. Today many parents who may not agree with the criteria used to assign movie ratings still forbid their children to see movies rated PG-13 or R; it is too hard for them to weigh the merits of each movie by themselves.
>
> Labeling organizations must choose vocabularies carefully to match the criteria that most people care about, but even so, no single vocabulary can serve everyone's needs. Labels concerned only with rating the level of sexual content at a site will be of no use to someone concerned about hate speech. And no labeling system is a full substitute for a thorough and thoughtful evaluation: movie reviews in a newspaper can be far more enlightening than any set of predefined codes."

3.39915533 Will the expense of labeling "flatten" speech by leaving non-commercial speech unlabeled, and hence invisible?

This is indeed a serious concern, explored in detail by Jonathan Weinberg in his law review article, Rating the Net <http://www.msen.com/~weinberg/rating.htm>. The following excerpt from Filtering the Internet <http://www.sciam.com/0397issue/0397resnick.html> acknowledges that materials of limited appeal may not reach even the audiences they would appeal to, but argues that labeling is merely a symptom rather than a cause of this underlying problem:

"Perhaps most troubling is the suggestion that any labeling system, no matter how well conceived and executed, will tend to stifle noncommercial communication. Labeling requires human time and energy; many sites of limited interest will probably go unlabeled. Because of safety concerns, some people will block access to materials that are unlabeled or whose labels are untrusted. For such people, the Internet will function more like broadcasting, providing access only to sites with sufficient mass-market appeal to merit the cost of labeling.

While lamentable, this problem is an inherent one that is not caused by labeling. In any medium, people tend to avoid the unknown when there are risks involved, and it is far easier to get information about material that is of wide interest than about items that appeal to a small audience."

3.399156 Filtering:

3.3991561 What is PICSRules?

PICSRules is a language for expressing filtering rules (profiles) that allow or block access to URLs based on PICS labels that describe those URLs. The purposes for a common profile-specification language are:

3.3991562 Sharing and installation of profiles:

Sophisticated profiles may be difficult for end-users to specify, even through well-crafted user interfaces. An organization can create a recommended profile for children of a certain age. Users who trust that organization can install the profile rather than specifying one from scratch. It will be easy to change the active profile on a single computer, or to carry a profile to a new computer.

3.3991563 Communication to agents, search engines, proxies, or other servers:

Servers of various kinds may wish to tailor their output to better meet users' preferences, as expressed in a profile. For example, a search service can return only links that

match a user's profile, which may specify criteria based on quality, privacy, age suitability, or the safety of downloadable code.

3.3991564 Portability betwen filtering products:

The same profile will work with any PICSRules-compatible product.

3.3991565 Does PICS make national firewalls easier to implement?

No, but an effective national firewall would make it possible for a government to impose PICS-based filtering rules (or non PICS-based filtering rules) on its citizens. A firewall partitions a network into two components and imposes rules about what information flow between the two components. The goal of a national firewall is to put all the computers in the country into one component, and all computers outside the country into the other component. This is difficult to do, especially if people deliberately try to find out connections (e.g., telephone lines) between computers inside the country and those outside the country. Given a successful partition, however, PICS could be used to implement the filtering rules for a firewall. In particular, the government could identify prohibited sites outside the country that people inside the country could not access; such a filtering could be implemented based on PICS-formatted labels or, without relying on PICS-compatible technology, with a simple list of prohibited URLs.

3.3991566 Does PICS make national firewalls easier to implement?

No. PICSRules can provide a way to express filtering preferences, but has no impact on the ability of a government to partition the computers inside a country from those outside the country.

3.3991567 Does PICS enable ISP compliance with government requirements that they prohibit access to specific URLs?

ISP compliance with government prohibition lists is already practical, even without PICS. It would also be possible to comply using PICS-based technologies. PICS does make it

easier for ISPs to comply with a government requirement to block access to sites labeled by non-governmental entities (including those that are self-labeled by the authors of the sites).

3.3991568 Does PICSRules enable ISP compliance with government requirements that they prohibit access to specific URLs?

Governments can make such requirements with or without PICSRules. PICSRules does make it possible for governments to precisely state filtering requirements, and perhaps simplify ISP compliance with changing government requirements, if the ISP implements a general interpreter for the PICSRules language.

3.39915691 Are proxy-server based implementations of PICS filters compatible with the principle of individual controls?

Yes. PICS enables mixing and matching of the five roles. In particular, a service provider could install and run filtering software on a proxy server, but allow individuals to choose what filtering rules will be executed for each account. AOL already offers a primitive version of this idea, not based on PICS; parents can turn the preset filtering rules on or off for each member of the family.

3.39915692 Are client based implementations of PICS filters usable only for individual controls?

No. Governments could require the use of filters on clients. The city of Boston, for example, requires public schools to install a client-based filtering product on all computers with Internet access, and requires public libraries to install a client-based filtering product on all computers designated for children.

3.39915693 Does my country have a right to filter what I see?

W3C leaves this question to the political and legal processes of each country. Some people argue that unrestricted access to information is a fundamental human rights question that transcends national sovereignty. W3C has not adopted that position.

3.39915694 Does my employer have a right to filter what I see?

W3C leaves this question to the political and legal processes of each country.

3.39916 PICS Technical Specifications:

3.399161 Completed Specifications for PICS-1.1

These are official W3C recommendations. They are stable. The normative specifications are in English. Translations <translations> of some of these are available.

Service descriptions: <http://www.w3.org/TR/REC-PICS-services> Specifies the format for describing a rating service's vocabulary and scales; analogous to a database schema.

Label format and distribution: <http://www.w3.org/TR/REC-PICS-labels> Specifies the format of labels and methods for distributing both self-labels and third-party labels.

PICSRules: <http://www.w3.org/TR/REC-PICSRules> Specifies an interchange format for filtering preferences, so that preferences can be easily installed or sent to search engines.

PICS Signed Labels (DSig) 1.0 Specification <http://www.w3.org/TR/REC-DSig-label/><http://www.w3.org/TR/PR-DSig-label> Specifies the syntax and semantics of digital signatures in PICS labels.

3.39917Lists of PICS-compatible products and services:

Technology Inventory <http://www.research.att.com/~lorrie/pubs/tech4kids/>. Lorrie Cranor and Paul Resnick. This inventory was first distributed at the December 1997 Internet On-line summit: Focus on Children. The on-line version was updated until the summer of 1999. It also lists some products and services that are not PICS-compatible.

The following resource lists are being maintained by members of the PICS developers' community. Contact the maintainer of each individual list with additional links. The maintainers have all agreed to be fast and fair in maintaining these lists (please send any unresolved complaints to pics-ask@w3.org).

- Client software <http://www.microsys.com/pics/> that reads PICS labels.

- HTTP servers <http://www1.raleigh.ibm.com/pics/servers.html> that distribute labels along with documents.
- Proxy servers <http://www.n2h2.com/pics/proxy_servers.html> that perform filtering based on PICSRules.
- Label bureaus <bureaus.htm>: HTTP servers that distribute third-party PICS labels through the PICS label bureau query protocol.
- Rating services <raters.htm>
- Search engine <http://www.xav.com/scripts/search/> that can use PICS labels in its selection criteria
- more information <http://www.getnetwise.org> "for families and caregivers" from GetNetWise <http://www.getnetwise.org/supporters.shtml>

3.3992 PORTABLE NETWORK GRAPHICS (PNG):

3.39921 Introduction:

PNG is an extensible file format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF and can also replace many common uses of TIFF. Indexed-color, grayscale, and truecolor images are supported, plus an optional alpha channel for transparency. Sample depths range from 1 to 16 bits per component (up to 48bit images for RGB, or 64bit for RGBA).

The PNG specification was issued as a *W3C Recommendation* on 1st October, 1996 . It is available in several formats:

- One big HTML file </TR/REC-png.html> (215k) or gzipped </TR/REC-png.html.gz> (68k)

- Multiple HTML files, </TR/REC-png-multi.html>

- Plain text </TR/REC-png.txt> (235k) or gzipped </TR/REC-png.txt.gz> (69k)

- PostScript </TR/REC-png.ps> (338k) or gzipped </TR/REC-png.ps.gz> (116k) or PDF </TR/REC-png.pdf> (306k)

This means it is a mature document that is considered to contribute towards realising the full potential of the Web. Viewers for PNG are available on many platforms; there are an increasing number of content creation tools available; and thus modern browsers implement support for it. The MIME type for PNG, approved in 14 October 1996.

3.3993 PLATFORM FOR PRIVACY PREFERENCES (P3P):

3.39931 Introduction:

The Platform for Privacy Preferences Project (P3P), developed by the World Wide Web Consortium, is emerging as an industry standard providing a simple, automated way for users to gain more control over the use of personal information on Web sites they visit. At its most basic level, P3P is a standardized set of multiple-choice questions, covering all the major aspects of a Web site's privacy policies. Taken together, they present a clear snapshot of how a site handles personal information about its users. P3P-enabled Web sites make this information available in a standard, machine-readable format. P3P enabled browsers can "read" this snapshot automatically and compare it to the consumer's own set of privacy preferences. P3P enhances user control by putting privacy policies where users can find them, in a form users can understand, and, most importantly, enables users to act on what they see.

The following resources provide information on P3P-enabling a web site:

- P3P Implementation Guide <http://p3ptoolbox.org/guide/> - why implement P3P, how does it work, preparing for implementation, creating P3P files

- P3P Deployment Guide </TR/p3pdeployment> - contains more technical details on deployment and server configuration than the Implementation Guide

- How to Create and Publish Your Company's P3P Policy (in 6 Easy Steps) <details.html> - a very brief overview

- P3PToolbox.org <http://www.p3ptoolbox.org/> -- features more information about P3P, P3P software, and upcoming P3P implementation workshops

- W3C P3P Validator <validator> -- to make sure your P3P enabled web site is working properly.

3.39932 Current Specifications, Working Drafts and Notes:

- **Platform for Privacy Preferences (P3P1.0) Recommendation W3C Recommendation** (Massimo Marchiori, editor) The final, normative specification of the Platform for Privacy Preferences (P3P). This document, along with its normative references, includes all the specification necessary for the implementation of interoperable P3P applications.
- **A P3P Preference Exchange Language 1.0 (APPEL1.0):** W3C Working Draft (Marc Langheinrich, editor). The latest public draft of the language for exchanging privacy preferences. This document complements the P3P1.0 specification [P3P10] by specifying a language for describing collections of preferences regarding P3P policies between P3P agents.
- **The Platform for Privacy Preferences 1.0 Deployment Guide:** W3C Note (Martin Presler-Marshall, editor). This is a guide to help site operators deploy the Platform for Privacy Preferences (P3P) on their site. It provides information on the tasks required, and gives guidance on how to best complete them.
- **A P3P Assurance Signature Profile:** W3C Note (Joseph Reagle, editor). This is not a normative specification. Instead, it captures the authors' thoughts on how applications might use the XML Signature specification to meet their requirements (defining signature semantics and algorithm profile); the example application is P3P. This is not a product or deliverable of any Working Group, nor does it reflect a consensus on how to use XML Signature's SignatureProperty. Instead this document

presents a possible use of SignatureProperty, as permitted (but not required) by the XML Signature specification, for further exploration/discussion.

- **An RDF Schema for P3P:** W3C Note (Brian McBride, Rigo Wenning and Lorrie Cranor, editors). This document describes an RDF Schema for P3P.

3.3994 QUALITY ASSURANCE ACTIVITY:

3.39941 Introduction:

W3C creates the technical specifications regarded by the Web community at large as Web standards. In order for these standards to permit full interoperability and access to all, it is very important that the quality of implementation be given as much attention as standards development. There has always been and still is a strong demand from the Web community, including end users, Web agencies, organizations, and software developers, for better support and better implementation of W3C specifications in both commercial and non-commercial products.

As of November 2002, W3C has published about 46 Recommendations. As our family of specifications gets more complex, their acceptance and deployment on the market becomes an ongoing challenge. Past experience with HTML, CSS and more recently SMIL, all implemented with various degrees of conformance by vendors, were strong incentives to start the QA Activity with due diligence.

The Quality Assurance (QA) Activity at W3C has a dual focus: to solidify and extend current quality practices, and to educate by sharing our understanding of coordination, certification, funding, and tracking of the quality of products and services related to W3C technologies. The mission of the QA Team is to improve the quality of W3C specification implementation in the field. In order to achieve that end, the QA Activity:

- Works on the quality of the specs themselves (e.g., to make sure they have a conformance section, a primer, clear text that is unambiguous for developers, good

layout, consistency between specifications, and in particular, that they are coordinated with the TAG </2001/07/19-tag>).

- Promotes the development of good validators, test tools and harnesses for implementors and end users.

- Thinks ahead in terms of what additional steps could be taken to achieve QA goals more efficiently, including certification, education, and communication.

W3C creates the technical specifications regarded by the Web Community at large as "Web standards", but to lead the Web to its full potential, it must ensure that its deliverables - W3C Recommendations <http://www.w3.org/TR/> - are implemented correctly. W3C has decided to take a new lead in improving the quality of implementation for W3C technologies. The Quality Assurance Activity </QA/Activity> gathers and formalizes QA efforts for the various languages and protocols.

3.39942 QA Activity has published two W3C Notes: (CHIPs and CUAP)

28 January 2003: The QA Activity has released two W3C Notes. CHIPs <http://www.w3.org/TR/2003/NOTE-chips-20030128/> is a set of good practices to improve implementations of HTTP and related standards as well as their use. It explains a few basic concepts, points out common mistakes and misbehaviors, and suggests "best practices". CUAP <http://www.w3.org/TR/2003/NOTE-cuap-20030128> explains some common mistakes in user agents due to incorrect or incomplete implementation of specifications, and suggests remedies. It also suggests some "good behavior" where specifications themselves do not specify any particular behavior (e.g., in the face of error conditions). This document is not a complete set of guidelines for good user agent behavior.

3.39943 Role of W3C:

Several W3C Working Groups, as well as individuals from the W3C Team and the Web community have started to assemble test suites, produce validation tools and follow

good QA practices. In addition, external organizations such as NIST and OASIS, and individuals have been active in the field of conformance and testing of W3C technologies, with varying degrees of W3C Working Group coordination.

These existing efforts are important and serve as a basis for future work as we move forward in the life cycle of our QA Activity at W3C. In order to be really effective, however, QA work for W3C technologies must be driven from inside W3C and must coordinate with all W3C Activities.

QA is absolutely necessary in order to ensure interoperability and usability and also to have consistency between the specifications W3C produces. The Web community, industry, and Members will benefit from the QA Activity as a guarantee of interoperable products, which is the core mission of W3C.

3.39944 Current Situation:

The QA Working Group is mainly working on a framework of documents to help W3C Working Groups and external bodies achieve quality with regards to our specifications. The list of The Seven Framework Documents <WG/> is available and updated regularly. Six parts of this framework were published as Working Drafts, three as Last Call:

- QA Framework: Introduction </TR/qaframe-intro> (Last Call)
- QA Framework: Operational Guidelines </TR/qaframe-ops> (Last Call)
- QA Framework: Operational Examples and Techniques <http://www.w3.org/QA/WG/qaframe-ops-extech>
- QA Framework: Specification Guidelines </TR/qaframe-spec> (Last Call)
- QA Framework: Specification Examples and Techniques </QA/WG/qaframe-spec-extech>

- QA Framework: Test Suite Guidelines </TR/qaframe-test> (Working Draft)

3.39945 The Future:

Test development is expected to be decentralized and done primarily in W3C Working Groups, with the QA Working Group monitoring the process to ensure consistency and timeliness.

The QA Working Group is working on the publication of new drafts for the entire QA Framework, and has has started to review materials and organization of other W3C Working Groups to improve its own material and to help Working Groups define a better QA strategy.

The QA Team and the QA Interest Group are working on resources and tools (like MUTAT, a test framework) to help people who want to promote best practices. The HTML Validator has been taken under the responsibility of the QA Team. A growing list of tutorials <http://www.w3.org/2002/03/tutorials> is collected. We have started to maintain liaisons with external user groups (like the WaSP) to improve education and outreach.

3.3995 METADATA:

3.39951 RESOURCE DESCRIPTION FRAMEWORK (RDF):

If the resource identified by a URL is unavailable, the URL is nearly useless. Reliability of URLs is currently subject to a number of single points of failure. As the number of users of the Web increases, and as the Web is increasingly used in mission-critical applications, it becomes clear that redundancy at all points is a requirement.

With each level of indirection and redundancy comes the possibility of version skew, forgery, and corruption. Information providers and consumers should be able to select from a number of authentication mechanisms to balance between integrity, speed, and convenience.

It is expected to be able to address robustness (reliability and authenticity) through the use of RDF.

3.399511 Introduction

The Resource Description Framework (RDF) integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax. The RDF specifications provide a lightweight ontology system to support the exchange of knowledge on the Web.

One of the major issues of the World Wide Web as it exists today is that it is really hard to automate any tasks which one has to perform on the web. So far, the web is mainly built as a forum for human interaction; because most web documents are written for human consumption, the only available form of searching on the web (for example) is to simply match words or sentences contained in documents. Anyone who has used a web search service like AltaVista or HotBot knows that typing in a few keywords and receiving a couple of thousand "hits" is not necessarily very useful. A lot of manual "weeding" of information has to happen after that; it may also happen that the keywords for which you are searching are not prominent in the relevant document itself.

A possible solution for the search problem - and for the general issue of letting automated "agents" roam the web performing useful tasks - is to provide a mechanism which allows a more precise description of things on the web. This, in turn, could elevate the status of the web from machine-readable to something we might call machine-understandable.

Metadata is "data about data" or specifically in our current context "data describing web resources." The distinction between "data" and "metadata" is not an absolute one; it is a distinction created primarily by a particular application ("one application's metadata is another application's data").

3.399512 Standardization Efforts at W3C:

One could say that the history of metadata at W3C begins with PICS or Platform for Internet Content Selection. PICS is a mechanism for communicating ratings of web pages from a server to clients; these ratings, or rating labels, contain information about the content of web pages: for example, whether a particular page contains a peer-reviewed research article, or was authored by an accredited researcher, or contains sex, nudity, violence, foul language etc. Instead of being a fixed set of criteria, PICS introduced a general mechanism for creating rating systems. Different organizations could rate content based on their own objectives and values, and users - for example, parents worried about their children's web usage - could set their browser to filter out any web pages not matching their own criteria. Development of PICS was motivated by the anticipation of restrictions on the Internet such as some recent US legislation (the Communications Decency Act and its subsequent overruling by the Federal Supreme Court).

PICS is a restricted metadata framework. It allows certain things to be expressed very precisely about web pages; in particular, PICS is useful when all the possible data values can be known in advance. The development of RDF as a general metadata framework - and in a way as a general knowledge representation mechanism for the web - was heavily inspired by PICS.

RDF - the Resource Description Framework, as our proposed mechanism is called - is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources. RDF metadata can be used in a variety of application areas; for example: in resource discovery to provide better search engine capabilities; in cataloging for describing the content and content relationships available at a particular Web site, page, or digital library; by intelligent software agents to facilitate knowledge sharing and exchange; in content rating; in describing collections of pages that represent a single logical "document"; for describing intellectual property rights of Web

pages, and in many others. RDF with digital signatures will be key to building the "Web of Trust" for electronic commerce, collaboration, and other applications.

RDF encourages the view of "metadata being data" by using XML (the eXtensible Markup Language) as its encoding syntax. The resources being described by RDF are, in general, anything that can be named via a URI (Uniform Resource Identifier). The broad goal of RDF is to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines the semantics of any application domain. The definition of the mechanism should be domain neutral, yet the mechanism should be suitable for describing information about any domain.

The recently published document about RDF introduces a model for representing metadata and one possible syntax for expressing and transporting this metadata in a manner that maximizes the interoperability of independently developed web servers and clients. This document is to be followed by others addressing issues such as how to define schemata (classes) for metadata, how to write queries, etc.

At the core, RDF data consists of nodes and attached attribute/value pairs. Nodes can be any web resources (pages, servers, basically anything for which you can give a URI), even other instances of metadata. Attributes are named properties of the nodes, and their values are either atomic (text strings, numbers, etc.) or other resources or metadata instances. In short, this mechanism allows us to build labeled directed graphs.

The essence of RDF is the model of nodes, attributes, and their values. In order to store instances of this model into files or to communicate these instances from one agent to another, we need a graph serialization syntax. The particular language we use is XML (XML being W3C's work-in-progress to define a richer Web syntax for a variety of applications). RDF and XML are complementary; there will be alternate ways to represent the same RDF data model, some more suitable for direct human authoring.

RDF in itself does not contain any predefined vocabularies for authoring metadata. We do, however, expect that standard vocabularies will emerge, after all this is a core requirement for large-scale interoperability. Some of the vocabularies in the foreseeable future are a PICS-like rating architecture, a digital library vocabulary (currently referred to as "Dublin Core"), and a vocabulary for expressing digital signatures. Anyone can design a new vocabulary, the only requirement for using it is that a designating URI is included in the metadata instances using this vocabulary. This use of URIs to name vocabularies is an important design feature of RDF: many previous metadata standardization efforts in other areas have foundered on the issue of establishing a central attribute registry. RDF permits a central registry but does not require one.

3.3991513 Future of Metadata on the Web:

The RDF working group - the W3C vehicle for crafting new standards - includes representatives from key companies and organizations: Netscape, Microsoft, IBM, Nokia, OCLC, etc. The interest from the large web browser vendors gives us hope that large scale deployment of tools which understand about RDF will take place; this in turn should lead to the widespread adoption of RDF on the web.

Once the web has been sufficiently "populated" with rich metadata, what can we expect? First, searching on the web will become easier as search engines have more information available, and thus searching can be more focused. Doors will also be opened for automated software agents to roam the web, looking for information for us or transacting business on our behalf. The web of today, the vast unstructured mass of information, may in the future be transformed into something more manageable - and thus something far more useful.

3.3991514 Dublin Core, Open Directory, and General Purpose Catalogs:

- Dublin Core Metadata Initiative

  The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a

broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices.

- OCLC Connexion (formally CORC - Cooperative Online Resource Catalog): CORC became a production service in July 2000; it started as a research project exploring the cooperative creation and sharing of metadata by libraries.
- Open Directory Project- The Open Directory Project is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors.
- RDF dumps of the are available. These dumps don't quite conform to the final RDF specification but rather to an earlier working draft.
- xmlTree - An index of XML content providers. The index is served in both RDF form and presented for human readability.
- DSpace - A newly developed digital repository created to capture, distribute and preserve the intellectual output of MIT.

  The dspace <../2000/01/sw/> component of Semantic Web Advanced Development aims to survey existing RDF data stores and examine effective techniques for storing complex metadata in a variety of systems.
- TAP - TAP's goals are to enable the Semantic Web by providing some simple tools that make the web a giant distributed Database. TAP is open source development effort by R.V. Guha (IBM) and Rob McCool (Stanford) which provide a set of protocols and conventions that create a coherent whole of independantly produced bits of information, and a simple API to navigate the graph. Local, independantly managed knowledge bases can be aggegated to form selected centers of knowledge useful for particular applications.

3.3991515 Perl Developers:

- The RDFStore Perl RDF API <http://rdfstore.sourceforge.net/> by Alberto Reggiori is a system for managing RDF models in Perl and includes a Perl version of the Stanford Java RDF API, the RDQL query language and persistent storage.

3.3991516 Python Developers:

- The Redfoot RDF framework by James Tauber and Dan Krech provides a system for building distributed data-driven web applications with RDF and Python. It includes an RDF database, query API, template language, module architecture, editor all with web interface, sample applications and the beginnings of P2P support.
- The W3C Semantic Web Area for Play by Tim Berners-Lee and Dan Connolly contains lots of small Python tools for RDF and beyond-RDF research tools including the Closed World Machine (CWM) data manipulator, rules processor and query system mostly using using the Notation 3 textual RDF syntax. Available under W3C open source license.
- The 4Suite 4RDF Python library provides open source tools for manipulating and querying RDF data, including inference capabilities.

3.3991517 C developers:

- The Redland RDF Application Framework library by Dave Beckett, Institute for Learning and Research Technology, University of Bristol is a portable C library that provides a high-level, object based interface for RDF allowing the model to be stored persistently, queried and manipulated. Includes the Raptor RDF Parser Toolkit for handling RDF/XML and other syntaxes and also provides Perl, Python, Tcl and Java interfaces. Available under open source / free software licenses (LGPL, GPL, MPL).
- RDFDB is a database that supports the RDF data model directly. It can load data from files or data can be inserted using the database API. It also supports an SQL-like query language. RDFDB is a very scalable and very fast triple store by R.V. Guha

3.3991518 Tcl/Tk developers

- The XWMF - An eXtensible Web Modeling Framework project by Alexander Block and Reinhold Klapsing contains various tools in Tcl including a parser, data model query tool and graphical editor. Available under open source and/or LGPL licenses.

3.39915191 PHP developers

- RAP RDF API for PHP by Chris Bizer. A pure PHP package for manipulating RDF models and parsing/serializing the RDF/XML syntax. See also the documentation and online demo. GPL License. Version 0.3 with new rdf:nodeID and rdf:datatype support announced 13 January 2003.

3.39915192 Related Technolologies:

Conceptual Graphs

- Corese(A COnceptual REsource Search Engine): The Corese platform implements an RDF/RDFS processor based on Conceptual Graphs (CG). Corese enables the processing of RDF Schemas and RDF statements within the CG formalism. The graph matching algorithm, called projection, enables to retrieve RDF statements according to a query and hence implements a search engine.

3.3996 TIMED-TEXT:

3.39961 Introduction:

The Timed-Text specification should covers all necessary aspects of timed text on the Web. Typical applications of timed text are the real time subtitling of foreign-language movies on the Web, captioning for people lacking audio devices or having hearing impairments, karaoke, scrolling news items or teleprompter applications.

The issue of developing an interoperable timed text format came up during the development of the SMIL 2.0 specification. Today, there are a number of incompatible formats for captioning, subtitling and other forms of timed text used on the Web. This means that when creating a SMIL presentation, the text portion often needs to be targetted to one particular

playback environment. This poses an issue for creating interoperable SMIL presentations. Moreover, the accessibility community relies heavily on captioning to make audiovisual content accessible to a hearing-impaired audience. The lack of an interoperable format adds a significant additional cost to the costs of captioning Web content, which are already high.

Timed Text will enrich the user experience for services involving timed text, and is seen as an important stimulus for instance in the usage of captioning and subtitling. The organizations willing to work on Timed Text include vendors of streaming multimedia technology, web browser companies, representatives of the accessibility community, caption content producers and consumer electronics companies.

3.39962 Players

Timed-Text 1.0

- None yet

3.39963 Authoring Tools

- None yet

3.39964 Demos

- None yet

3.3997 VALIDATORS:

3.39971 Log Validator:

Log Validator is a web server log analysis tool which finds the N most popular documents matching a particular criteria. Thanks to a modular, extensible design, the criteria can be chosen and modified arbitrarily. The Log Validator was first written with Validation (HTML, etc.) in mind : it can thus help web content managers find and fix the most frequently accessed invalid documents on their Web site, acting as a comprehensive validation tool.

The first HTML validator service has been a way to check the validity of one's webpage with regards to web standards (HTML, CSS...). Other services, like HTML Tidy allows to (semi-)automatically fix invalid documents. This tool is here to make your life as a webmaster, web designer, web developer even easier, by telling you which documents you should fix in priority. It has first been developed by Gerald Oskoboiny as an internal W3C tool (yes, even at W3C we create invalid HTML sometimes) to check the HTML validity of the webpages on the W3C website.

In 2002, the Quality Assurance team at W3C decided to re-write it as a modular, portable, and easy-to-use tool for webmasters. This tool takes a web server's last logs and processes it through *validation modules*. Those *validation modules* check the most popular documents' validity for a certain technology.

The (X)HTML validation module, for example, helps you find, among the most popular pages on your site, which are invalid, and thus tell you which (invalid) pages you should fix first. This is a step-by-step process, you can set up this tool to run every week, and painlessly fix only a few documents at the time. Eventually, you will have fixed your whole site, or at least the most important parts of it.

3.39972 MUTAT:

MUTAT is a simple cgi script to demonstrate possible uses of EARL and also use of RDF to configure a QA helper application. QA experts have come up with a variety of solutions to handle entry and retrieval of the data they need - sometimes little more than a paper and pencil. The idea behind MUTAT was to demonstrate automating the process for human-centered testing, and output the results in EARL. The script presents the user with a series of questions: first a page of fill-in information on their identity and testing environment, and then the actual tests. The tests can be formatted either as a single page of text-based questions and links or a succession of framed HTML pages, with form inputs for the results. Both the initial information questions and the test questions are configured using an RDF file. There is also a feature that allows filtering of test questions.

3.39973 TODO:

- Update to new version of EARL as proposed by the working group
- Follow current discussion on creation of a public EARL store
- Accept XML RDF
- Error-check on user input
- Allow for completion of partial reports (should be possible with working EARL store)

You can go straight to the bare front page of the testing tool with this link <bin/mutat>, but that probably won't be very interesting without the URI of a file containing tests. There are two existing examples of test files that show off different features in the script. The links show how you can use GET-style URIs to set values in the test and customize the opening page.

3.3998 VOICE BROWSER:

3.39981 Introduction:

W3C is working to expand access to the Web to allow people to interact via key-pads, spoken commands, listening to prerecorded speech, synthetic speech and music. This will allow any telephone to be used to access appropriately designed Web-based services, and will be a boon to people with visual impairments or needing Web access while keeping their hands and eyes free for other things. It will also allow effective interaction with display-based Web content in the cases where the mouse and keyboard may be missing or inconvenient.

To fulfill this goal, the W3C Voice Browser Working Group is defining a suite of markup languages covering dialog, speech synthesis, speech recognition, call control and other aspects of interactive voice response applications. Specifications such as the Speech Synthesis Markup Language, Speech Recognition Grammar Specification, and Call Control XML are core technologies for describing speech synthesis, recognition grammars, and call

control constructs respectively. VoiceXML is a dialog markup language that leverages the other specifications for creating dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key (touch tone) input, recording of spoken input, telephony, and mixed initiative conversations.

These specifications bring the advantages of web-based development and content delivery to interactive voice response applications. Further work is anticipated on enabling their use with other W3C markup languages such as XHTML, XForms and SMIL. This will be done in conjunction with other W3C Working Groups, including the Multimodal Interaction Activity.

Some possible applications include:

- Accessing business information, including the corporate "front desk" asking callers who or what they want, automated telephone ordering services, support desks, order tracking, airline arrival and departure information, cinema and theater booking services, and home banking services.

- Accessing public information, including community information such as weather, traffic conditions, school closures, directions and events; local, national and international news; national and international stock market information; and business and e-commerce transactions.

- Accessing personal information, including calendars, address and telephone lists, to-do lists, shopping lists, and calorie counters.
- Assisting the user to communicate with other people via sending and receiving voice-mail and email messages.

3.39982 Current Situation:

W3C's work on voice browsers originally started in the context of making the Web accessible to more of us, more of the time. In October 1998, W3C organized a workshop on

"Voice Browsers". The workshop brought together people involved in developing voice browsers for accessing Web based services. The workshop concluded that the time was ripe for W3C to bring together interested parties to collaborate on the development of joint specifications for voice browsers. As a response, an activity proposal and charter were written to establish a W3C "Voice Browser" Activity.

The organizations currently participating in the Voice Browser Working Group are:

BeVocal, Canon, Comverse, France Telecom, Genesys, HeyAnita, Hitachi, HP, IBM, Intel, IWA/HWG, Loquendo, Microsoft, MITRE, Mitsubishi Electric, Motorola, Nokia, Nortel Networks, Nuance, PipeBeach, SAP, Scansoft, Snowshore Networks, SpeechWorks, Sun Microsystems, Syntellect, Tellme Networks, Unisys, Verascape, Vocalocity, VoiceGenie, Voxeo, and Voxpilot

3.39983 Work under development:

The major work done for the development by the Voice Browser working group is discussed herewith. The suite of specifications is known as the W3C Speech Interface Framework.

The top priority work items cover dialog (VoiceXML), speech recognition grammar, speech synthesis, semantic interpretation, and call control. The lower priority items cover: pronunciation lexicon, stochastic grammars (N-Grams) and voice browser interoperation.

In early 2003, it is started collecting requirements for future versions of the dialog markup language. Against this an interim release named VoiceXML 2.1, based upon a small set of extensions to VoiceXML 2.0, that have been widely implemented throughout the industry. Now work is going on closely with the Multimodal Interaction activity to provide support for speech within multimodal applications. This will fold into the next major version of the dialog markup language, which we have code-named "V3".

3.399831 VoiceXML 2.0

VoiceXML 2.0 is designed based upon extensive industry experience for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations.

Based upon a small set of widely implemented extensions to VoiceXML 2.0, an interim version of the dialog markup language called **VoiceXML 2.1**. These features will help developers build even more powerful, maintainable and portable voice-activated services, with complete backward compatibility with the VoiceXML 2.0 specification. It is expected to publish VoiceXML 2.1 as a small specification that describes the extensions to 2.0.

3.399832 Speech Recognition Grammar (SRGS)

The speech recognition grammar specification covers both speech and DTMF (touch-tone) input. DTMF is valuable in noisy conditions or when the social context makes it awkward to speak. Grammars can be specified in either an XML or an equivalent augmented BNF (ABNF) syntax, which some authors may find easier to deal with. Speech recognition is an inherently uncertain process. Some speech engines may be able to ignore "um's" and "aah's", and to perform partial matches. Recognizers may report confidence values. If the utterance has several possible parses, the recognizer may be able to report the most likely alternatives (n-best results).

3.399833 Speech Synthesis (SSML)

The Speech Synthesis specification defines a markup language for prompting users via a combination of prerecorded speech, synthetic speech and music. You can select voice characteristics (name, gender and age) and the speed, volume, pitch, and emphasis. There is also provision for overriding the synthesis engine's default pronunciation.

The Voice Browser working group is collaborating with the CSS working group to develop a CSS3 module for speech synthesis based upon SSML for use in rendering XML

documents to speech. This is intended to replace the aural cascading style sheet properties in CSS2.

3.399834 Semantic Interpretation

The semantic interpretation specification describes annotations to grammar rules for extracting the semantic results from recognition. The annotations are expressed in a syntax based upon a subset of ECMAScript, and when evaluated, yield a result represented either as XML or as a value that can be held in an ECMAScript variable. The target for the XML output is EMMA (Extensible Multimodal Annotation Markup Language) which is being developed in the Multimodal Interaction activity.

3.399835 Call Control (CCXML)

W3C is working on markup to enable fine-grained control of speech (signal processing) resources and telephony resources in a VoiceXML telephony platform. The scope of these language features is for controlling resources in a platform on the network edge, not for building network-based call processing applications in a telephone switching system, or for controlling an entire telecom network. These components are designed to integrate naturally with existing language elements for defining applications which run in a voice browser framework. This will enable application developers to use markup to perform call screening, whisper call waiting, call transfer, and more. Users can be offered the ability to place outbound calls, conditionally answer calls, and to initiate or receive outbound communications such as another call.

3.39984 Future work on dialog markup:

The purpose of the next major version of the dialog markup language (code named "V3") is to provide powerful dialog capabilities which can be used to build advanced speech applications, and to provide these dialog capabilities in a form which can be easily and cleanly integrated with other W3C languages. For example, the Multimodal Interaction Activity would be able to combine this dialog language with markup languages for other modalities to build multimodal applications. In comparison with VoiceXML 2.0, the

language will provide improved dialog functionality, greater flexibility, and be modularized so as to allow embedding in languages such XHTML, SMIL and SVG.

Work started in early 2003 on collecting detailed requirements for this dialog markup language. The requirements will be drawn from sources including deferred change requests on VoiceXML 2.0, other activities within the Voice Browser group (especially Call Control), external contributions such as SALT 1.0 and XHTML+Voice Profiles, and other interested working groups within W3C especially Multimodal, XHTML, and WAI. The requirements are expected to be published in the 3rd quarter of 2003. It is intended that the first working draft of this dialog language will be published in the first quarter of 2004.

3.39991 WEB COMPUTER GRAPHICS METAFILE:

3.399911 Introduction:

CGM (Computer Graphics Metafile) has been an ISO standard for vector and composite vector/raster picture definition since 1987. CGM has a significant following in technical illustration, interactive electronic documentation, geophysical data visualization, amongst other application areas and is widely used in the fields of automotive engineering, aeronatics, and the defence industry.

WebCGM is a profile for the effective application of CGM in Web electronic documents. WebCGM has been a joint effort of the CGM Open Consortium, in collaboration with W3C staff and supported by the European Commission Esprit project. It represents an important interoperability agreement amongst major users and implementors of CGM, and thereby unifies current diverse approaches to CGM utilization in Web document applications. WebCGM's clear and unambiguous conformance requirements will enhance interoperability of implementations, and it should be possible to leverage existing CGM validation tools, test suites, and the product certification testing services for application to WebCGM .

While WebCGM is a binary file format and is not "stylable", nevertheless WebCGM follows published W3C requirements for a scalable graphics format where such are

applicable. The design criteria for the graphical content of WebCGM aimed at a balance between graphical expressive power on the one hand, and simplicity and implementability on the other. A small but powerful set of metadata elements is standardized in WebCGM, to support the functionalities of: hyperlinking and document navigation; picture structuring and layering; and, search and query on WebCGM picture content.

3.399912 Status:

The WebCGM Profile specification was issued as a *W3C Recommendation* on 21st January 1999. This means it is a mature document that is considered to contribute towards realising the full potential of the Web. Viewers for CGM are available on many platforms and are being adapted to support the WebCGM Profile specification.

3.399913 MIME type:

CGM has been a registered MIME type since 1995. The MIME type for CGM is unusual in that it requires two parameters - the CGM version, and the CGM Profile in use. This is because, without Profiles, it is very difficult to achieve interoperability with CGM (which is why W3C issued a WebCGM Profile). It also means that the MIME type does not need to be re-registered whenever there is a new profile.

3.39992 CRITICAL ISSUES OF WEB-ENABLED TECHNOLOGIES:

3.399921 Introduction:

The Web-enabled technologies are not free of associated risks and controversies. Like many other emerging technologies, this technology has its share of associated problems and limitations. In order to clearly understand the total potentials of these technologies, one must also assess the limitations, stipulations and provisions of these technologies in modern organizations. Some of the many issues, problems and limitations of the Web-enabled technologies are given below.

3.399922 Bandwidth restrictions and latency:

A large percentage of the Web users run low-speed modems 56K, which in reality cause considerable delays in obtaining Web-based materials when the corresponding downloads incorporate images animation and audio (Berghel, 1996; Pitkow, 1996). A recent study by a popular server revealed that about one in five users was connecting with graphics turned off to eliminate the annoying latency of loading Web pages (Fox and Brewar, 1996). Latency issues are also being experienced with some of the more popular Web documents. In this case, the slowness relates to the number of requests an individual Web server can handle at once (Roush, 1995). Since the Web has evolved into a multimedia intensive tool, gridlock has become an even bigger problem than for the other part of the Internet.

3.399923 Cyberloafing:

Surfing the Internet, wasting time and accessing inappropriate materials are the primary concerns, which have been labeled as cyberloafing (Prawitt *et al.*, 1997). Studies show that, once users become more familiar with the Web, the cyberloafing practice becomes a common phenomenon (Hills, 1997; Frook, 1997). Cyberloafing can also take place in a different form, where users receive unsolicited messages about all kinds of decent and indecent offers. In this case the user is not searching sites to explore; however, in the act of reading the unsolicited e-mail message they can be tempted to explore inappropriate materials.

In the latter case when the cyberloafing takes place at work by an employee of an organization, besides the productivity lost, there is also the liability concern associated with cyberloafing, if the act involves the downloading of indecent materials. These actions can create a company liability that potentially involves allegations of "harassment", "free speech", "privacy", "jurisdiction" and even "copyright infringement" (Sampson, 1997).

3.399924 Equity:

Some argue that the Web will bring forth a better democracy within the USA by returning the power to the people (Meeks, 1997). This may not come to pass if the issues of

equity and demographic trends are poorly addressed. According to Pitkow (1996), of the users joining the Web, their estimated median income ($64,700 annually) is well above the national median of $36,950 as estimated by the 1993 US Census and they are predominantly male (70 per cent). Whether the explanations for the lack of utilization by some groups lie in the issue of availability, affordability or usability remains a topic for additional research. However, all statistics clearly indicate that this technology is not equally utilized by all classes of society in the USA as well as other countries.

3.399925 Exposure points:

As more companies utilizing Web-enabled technologies incorporate the ability for remote access to their computer systems by their employees, there is a higher risk of information exposure (Prawitt *et al.*, 1997). In other words, these emerging exposure points are inroads which can lead to sloppy data entry into the systems, as well as savvy hackers breaking into the system, where inadequate control measures might not be applied at every exposure point.

3.399926 Flooding of the Web with content for content's sake:

With the ease of access to the Internet and the availability of access to Web development tools, there is an abundance of slick and costly WebPages on the Internet. Many of these WebPages include information that is not helpful to their viewers. They are merely on the Internet so the individuals or company that owns the site can claim that they have a Web site. In recent years, many companies have been beginning to view content for content's sake as a wasteful exercise and instead are beginning to understand that the role of the Web site is to facilitate business processes (Gardner, 1997).

3.399927 Inadequate search facilities on the WWW:

One of the important issues of the Internet is that of inadequate search facilities with the lack of a high level query language for locating, filtering and presenting WWW information (Foo and Lim, 1997). Some search engines search the document headers, some look for the document themselves, while others look for indices or directories. As a result,

one can conclude that much of the information on the Web is presented in a dynamic and somewhat chaotic fashion. In a recent survey of Web users, 34.5 per cent of the participants were not able to locate a site which was known to exist and 23.7 per cent were not able to figure out how to return to a site that they had previously visited (Pitkow, 1996).

3.399928 Maintainability and integrity of data:

The task of keeping up with commercial WebPages by maintaining the latest information is considered to be a costly issue facing many organizations. As a company's Web site becomes more elaborate and complex, the task of maintaining and validating information included in their Web sites becomes much more costly and complex too. Ultimately, it will reach a point where maintaining and ensuring the accuracy of information becomes difficult (Foo and Lim, 1997). Inaccurate and out-of-date information included in the Web site can contribute, in part, to decisions being made by the user of the information that are based on data that are either inaccurate or outdated which can harm the organization's business processes.

3.3999291 Security:

The issue of Web security is considered as being among the most important challenges of many organizations. Many security experts believe that the existing layers of security are considered to be inadequate and in some cases fragile (Hodges, 1997). It is important to note that security is a broad term. In some instances the term security is related to privacy, while in other contexts the term refers to the integrity of data (Grimshaw, 1997). Security issues can be better understood by examining the concrete examples of security threats and risks.

3.3999292 E-mail risks:

Berghel (1997) and Prawitt *et al.* (1997) identify several unique risks related to e-mail, including volume levels that overload systems, junk mail, mail bombs, "flaming" or flooding a user with messages, interception and unauthorized reading of electronic mail, and improper representations by employees.

3.3999293 False store fronts:

A false store front risk is presented when a hacker sets up a Web page that looks legitimate for business but uses the site to gather credit card numbers, account numbers or other confidential information from unsuspecting consumers (Bhimani, 1996; Prawitt et al., 1997), after which, the "business" disappears and the information obtained is utilized for unauthorized transactions.

3.3999294 Industrial espionage:

There is a growing concern that the Web requires additional methods to secure the confidential data accessible on the Web against interception and decryption by unauthorized users (Roush, 1995). "Sensitive about anything that touches their legacy applications and custom-built accounting and inventory tools that run the business side of the corporation, most companies have tiptoed carefully into so-called Webification" (Higgins, 1997). Most of the attacks launched at industry systems take advantage of simple holes largely attributed to misconfigured systems, poorly written software, mismanaged systems, or user neglect (Bhimani, 1996).

3.3999295 Information vandalism

Vandalism in this context is the unauthorized modification of data that are available on the Web. Often this takes the form of "graffiti" placed in the text of a home page, which are unauthorized, often embarrassing and potentially harmful (Prawitt *et al.*, 1997). Another form is described by Bhimani (1996), whereby the contents of certain transactions are modified, such as the payee of an electronic check or the amount of a bank account transfer.

3.3999296 ISP linkage alterations:

Internet Service Providers (ISP) provide access to the Internet via the maintenance of a domain name server, which provides a direct translation of a WWW address into an Internet address. If a link established within a WebPage is altered, then the user may be

erroneously linked to an inappropriate, potentially embarrassing location within the Web (Prawitt *et al.*, 1997).

3.3999297 Viruses:

With the increasing number of networked computers, the ability of a developer to place a virus within any number of programs and have that virus become widespread to all who download, open or execute the program or file is great (Prawitt *et al.*, 1997).

3.3999298 Webware:

Many systems allow software developers to attach programs which are executed upon access to a WebPage (Felton, 1997). This software is termed Webware. "Simply visiting a WebPage may cause you to unknowingly download and run a program written by someone you don't know or don't trust" (Felton, 1997).

With the advent of Electronic Commerce (EC) and the overwhelming interest in utilization of this technology for modern commerce, there are many challenges presented by the security issues and risks. Although there are perceived issues with security, especially related to EC, there is still substantial interest in utilizing the Web technology for EC (Liu *et al.*, 1997).

3.39992991 System incompatibilities:

The issue of system incompatibilities has been dominant during the past several years. In many cases, cross-platform compatibility is not always available in all of the emerging technologies being developed which, can result in difficulty when trying to make them function in unison (Prawitt *et al.*, 1997).

3.39992992 Unauthorized use of computer resources:

Today's emerging interconnectivity technologies have presented opportunities for computer misuse which were not previously possible (Prawitt et al., 1997). The Boeing corporation recently has begun reviewing the issue of URL filtering of objectionable material

(Frook, 1997). According to the Computer Fraud and Abuse Act, computer usage in excess of one's level of authorization can result in personal liabilities for any harm caused (Sampson, 1997). Many companies are debating the best solution to this issue. In response, Boeing has decided that "restricting site access is a cumbersome management process, but unrestricted access to public Web sites could open the company up to legal issues" (Frook, 1997).

3.39992993 User ignorance and perceptions:

The lack of adequate understanding of the Internet and its usage and risks has been a contributing factor in maintaining secure systems. Modern information systems comprise many different components of distributed hardware, software and data maintained on different locations by different systems. According to Prawitt *et al.* (1997), while it is becoming increasingly critical for users to exercise sound control practices, most are not adequately trained to do so.

3.39992994 Web performance tracking:

With the explosive growth of Web applications, services, traffic volumes and contents, a management void has been created. If the performance and availability of the Web services are not managed and information cannot be accessed quickly, it is likely that the user will jump to a competitor's site, which results in the loss of business McConnell (1997) states as "To achieve peak performance, the IT department must harmonize many critical elements, including the transport network and its Featured Sites service levels, if any, Web server hardware and software and information content."

3.39992995 Low overhead e-payment facilities:

Low overhead e-payment facilities, micropayments, are needed as a service on the WWW, so that advertising is no longer necessary to cover the costs of running services and so that the content providers can sell information in the same fashion as the purchase of newspapers or a single song (Machlis, 1998). The process typically works, whereby an account is opened with a micropayment system and the software required is downloaded to

work with the user's browser. Digital Equipment Corporation has developed a system that "will eliminate minimum purchase requirements of ten to 25 cents now imposed by other electronic payment methods, allowing users to buy and sell information profitability down to fractions of a cent".

3.39992996 Failure to adhere to standards:

Failure of primary companies and their products to adhere to the standards that exist (e.g. HTML and JavaScript) is an issue that was ranked highly by the expert panel on the study. The primary importance of this issue lies in the fact that many companies, which have monopolistic power, such as Microsoft, do not abide by those standards that exist in the industry. Their lack of following the industry standards can be viewed as an attempt in creating a new "standard" with their products.

3.39992997 Unsolicited e-mail (spamming):

Spamming occurs when an endless stream of mail is received, which can overflow the user's mailbox and can even choke the user's system. In recent years, with the easy to obtain free e-mail addresses from many different sources such as "Hotmail", "Juno", and "Netscape", the act of forwarding unsolicited e-mail messages has reached a crisis level for many users everyday. Users all around the world receive unsolicited e-mail messages for promoting products or services.

3.39992998 Use of metadata:

The World Wide Web currently has a huge amount of data with practically no classification information and this makes it extremely difficult to handle data/information effectively (Marchiori, 1998). Many systems can support knowledge management by establishing a metadata - information about information - standard, so that users of data can obtain the raw materials that enable them to capture, store and share knowledge that is gathered from different sources (Phillips, 1995). This task can be accomplished by adding to Web objects a metadata classification which will assist search engines and Web-based digital

libraries to properly classify and structure the information on the WWW (Baer, 1996; Marchiori, 1998).

3.399929991 Ensure a continued global body:

Many users of the WWW are concerned that the body of knowledge created on the Internet consists of some kind of global understanding to which users from all over the world can relate. This task has been assigned to the World Wide Web Consortium (W3C) to accomplish. W3C is a global body that was founded to lead the WWW to its full potential by developing common protocols that promote its evolution and ensure its interoperability. The primary services offered by W3C to users and developers consist of:

- acting as a depository of information about the World Wide Web;
- providing reference code implementations to embody and promote standards; and
- providing various prototype and sample applications to demonstrate use of new technology.

3.399929992 Privacy and confidentiality agreements:

The privacy and confidentiality agreements issue entails an aspect of the security issue in that it is a violation of users' privacy. This issue addresses the dilemma of individual right to privacy and the sharing of confidential information about people in society. With the technology advances of the past two decades, many users believe that more information about their lives is now shared with others through the use of the Internet. Despite many existing laws with regard to the "right to privacy" of users of the Internet, everyday there are many cases of the violation of users' privacy and confidentiality, where information that should not be shared by others is passed throughout the Webs of this modern technology.

3.399929993 Global laws for Net crimes:

Despite the fact that there is a global perception that crimes and criminals should be punished, there is considerable confusion regarding what is a criminal act in one society in comparison with another. Global laws for Net crimes are considered to be a complex issue that is anticipated to remain unresolved for a long time based on the current dilemma of establishing national laws regarding Internet activities (Rose, 1996; Weston, 1996; Charlesworth, 1997).

3.399929994 Required labeling of sites:

With millions of Web sites in existence and millions added constantly, there is a concern about how to differentiate Web sites from one another regarding their contents. This particular issue can be considered an offset of two other critical issues discussed in this chapter, "inadequate search facilities" and "global laws". Supporters of this issue claim that by labeling sites search engines can provide more effective and efficient search processes, and also labeling will assist the enforcement of any global laws related to Net crimes. This issue becomes valid after having global laws in place to deal with Net crimes.

3.399929995 System utilization:

The issue of system utilization deals with the overall question of what functionality or information sharing is best served on the World Wide Web. During the past decade, many users have seen the transformation of this technology into a business tool, where businesses all over the world can conduct their commerce through this medium. Many users question what should be the overall functionality of the Internet in the future, as this becomes more acceptable as a common medium for communication purposes.

3.399929996 Expressibility of HTML:

The expressibility of HTML issue is primarily concerned with the ability of the user to create documents that contain complex layouts. This is very important for the functionality of the Web because of its usefulness in presenting information with all its characteristics and potentials. As more users rely on the use of this technology to share information with other users, the role of HTML or other tools will become more recognized. These tools should

allow users at all levels to create documents that contain the full picture and are not limited by the shortcomings of the tool that they used.

3.399929997 Lack of standardized vector graphics:

This issue deals with the lack of incorporation of vector graphics in Web design. The adoption of vector graphics in Web design would enable programmers to present better user interfaces to Web applications - "Vector graphics scale easier, download faster and print better than their bit-mapped graphics counterparts GIF and JPEG" (Walsh, 1998). Standardization in this facet of the Web is at an early stage, whereby a couple of proposals have been placed before the W3C for consideration (Walsh, 1998).

3.399929998 Hype:

Web sites are effective if used imaginatively and intelligently. Many firms boast that their Web sites are showcases for the firm's goods and services. Few, however, are very effective at serving the cause of the firm's betterment. Despite this reality, there is a considerable degree of hype among organizations and their Web designers that their Web sites should consist of more whistles and bells in order to compete with their rivals, as well as to attract more customers.

3.399929999Access appliances that avoid computer software management:

This particular issue was added by one of the expert panel members to the list of critical issues of this study during round one. (Other panel members questioned the real meaning of this issue but, regrettably, the expert who suggested the issue did not provide any clarification or further responses after the first round. At the time of writing, there is no clear explanation concerning this issue.

CHAPTER 4: DIGITIZATION

4.1 INTRODUCTION:

Technical Standards and Guidelines during digitization are the essential issues of the digitization process that should be taken care during the planning stages and discuss techniques for creating digital files that will conform to the guidelines. There may be valid institutional reasons for following or discarding different aspects especially in relation to the handling of original materials that may make certain processes unsuitable for that class of material. The fundamental issues associated with the digitization process are as follows:

4.2 KNOWLEDGE OF ORIGINAL DOCUMENTS:

Having a good knowledge of the contents of the collections that are intended to be digitized will make it much easier to decide on processes and techniques for converting the originals to digital form. The physical processes required to create a digitized version of an original item depend on many factors, including:

- The format of the original - is it printed text, photographic material, video, audio etc.?
- The condition of the original - will it stands up to automated procedures (if used), will conservation be required before scanning?
- The size of the original
- The color content of the original and whether those color is important.

For paper and photographic originals, issues to consider include the following:

4.21 Photographic media (transparencies, prints, and negatives)

- What size are the originals, are they all the same size? It makes for a smoother workflow if items of a similar size are grouped together.
- What proportion of the items have color content? Is it important to capture the color?

- What condition are they in, for example, are they dirty from heavy use? If they are dirty a better scan will be achieved if the items can be cleaned first.
- What format are they in? Slides in sleeves or strips will take longer to prepare for scanning and may cost more if a bureau is scanning them. Glass negatives are prone to breakage and require careful handling.
- Are the photographs flat or have they bowed? Bowed originals cause difficulties with focus and may need weighting down.
- What is the quality of the original? A bad original (i.e. out of focus) will not be improved by scanning.

4.22 Paper media:

- What size are the pages, are all items the same size?
- What general condition is the material in? Pristine pages will produce a better result and the scanning process may be able to be automated. Any damage in an original may be exacerbated by the scanning procedure.
- Can books that are bound be stripped to loose pages for scanning? Scanning from bound volumes is more complex and therefore expensive than from loose pages.
- Is there any artwork? - Is it black and white or color photographs or line art? Color scanning is generally more complex and resource intensive.
- Is the text size particularly small or large? Very small text may need a higher resolution to extract the information.
- Objects require a different approach. Artifacts, art works and sculptures cannot generally be successfully scanned using the techniques available for 'flat' media such as photographs. It will therefore be necessary to use photography, either traditional or digital, to get an image of the original.

4.3 DIGITIZATION: A TECHNICAL OVERVIEW:

Digital preservation issues must be observed when producing digital content. A good baseline for creating a digital file that will be long-lasting and would be scan once for all purposes, this means that all the complex and expensive preparation work will only need to be done once. The project should consider the value in creating a fully documented high-quality 'digital master file' from which all other versions (e.g. compressed versions for accessing via the Web) can be derived. This 'digital master file' should be created at the highest suitable resolution and bit depth that is both affordable and practical. This master file then becomes the source for every other version of that item that the project will require, such as Web surrogates, versions for high quality printing and so on. The 'digital master file' will become an archive version of the data - it remains as pure a representation of the original as possible. Ideally more than one copy should be stored on more than one media type and in more than one geographical location, thus providing a degree of protection against data corruption, media failure and physical damage to equipment. 'Surrogate' or 'access' versions of the digitized item can be created from the digital master file using image manipulation software such as Adobe Photoshop or Paintshop Pro.

4.4 RESOLUTION AND BIT-DEPTH:

Resolution is usually expressed in dots per inch (DPI) and relates to the density of information that is captured by the scanning equipment. Broadly speaking, the higher the DPI the more detail is being captured. The amount of resolution required to get a useful image of an item is determined by the size of the original, the amount of detail in the original and the eventual use for the data. For example, a 35mm transparency will require a higher DPI than a 5x4 print because it is smaller and more detailed. An A4 sized modern printed document that is intended to be processed into a searchable text will need less resolution than a similar sized photographic original. There are also upward limits on resolution - file size is one (increasing resolution will increase the file size) and another is preventing the capture of extraneous information. For example, postcards are often printed on poor quality paper and if they are scanned at too high a resolution the texture

of the paper will be captured and can obscure the content. There is also a point where putting more resolution into the capture process will no longer add value to the information content of the digital output. Suitable resolutions for digital master files for various media types are discussed in the HEDS Matrix and the JIDI Feasibility Study contains a useful table of baseline standards of minimum values of resolutions according to original material type. Bit-depth relates to the level of color that will be captured. A 'bit' is the binary digit that represents the tonal value of the pixel. As an overview, a 1-bit image is black and white (the pixel has 1 bit and is therefore black or white with no shades in between), an 8-bit image has 256 shades of either grey or color (28 = 256 shades), and a 24-bit image has millions of shades of color (224 = 16,777,216 shades). A detailed discussion of resolution, binary and bit depth can be found on TASI's Web pages and a good basic guide to color capture can also be found on the EPIcentre Web pages.

4.5 SCANNING EQUIPMENT:

Digitization equipment can be separated into 'contact' and 'no-contact'. 'Contact' equipment, i.e. flatbed scanners, requires that the original be flat against the scan bed to get a scanned image. This approach will only work if the original is flat or can be pressed flat without damage to it. No-contact equipment includes overhead scanners or book scanners and digital cameras that are able to obtain a digital image with the bare minimum of contact with the original. The equipment for scanning the originals will depend largely on the characteristics of the collection. In general terms, photographic materials are usually scanned on a flatbed or a transparency scanner while bound volumes and oversized flat materials such as maps and plans require a digital camera or an overhead scanner. The Feasibility Study for the JIDI project gives information about the type of equipment that is most suitable for broad groups of media types. If you have a mixed media collection then it may not be possible to use one scanner for everything. A flatbed that is ideal for high speed, high volume paper scanning may not be capable of the resolution required for high quality scans of transparencies. A digital camera studio set-up will be overkill for loose-leaf paper scanning and for most general photographic materials. Generally, it should make sure that requirements match the capability of the

scanner. Look carefully at the resolution that the scanner is capable of, the scanner will often be listed with a maximum optical resolution and an interpolated or software resolution. The optical resolution is the figure to look for - interpolated resolution uses software to 'guess' the values of pixels that are between those that the scanner can optically register. Interpolation should be avoided in an archive-quality scanning exercise. Where resolution is listed as, for example, 600x1200 DPI the maximum optical resolution will be 600. The dynamic range of the scanner is important - it describes the tonal density of the information that the scanner will be able to capture and generally speaking the higher this is the better, particularly for dense originals such as photographic prints and transparencies. A good flatbed scanner is often the keystone to a scanning unit. Production level flatbed scanners usually have either an A4 or an A3 sized scanning area. Larger ones are available but are specialist equipment and therefore rather expensive. In order to choose a flatbed it is required to know the size of the originals, whether they are reflective (i.e. light is bounced off them to capture the image, as in photographic prints) or transmissive (light is passed through the original to capture the image, as in transparencies), the resolution and bit depth to be capture and the volume of the work to be done. The software that runs the scanner is also important. It should be straightforward to use and an ability to run batch scans will save time as the scan bed can be loaded with originals and more or less left to get on with it. The Digital Eyes Web site lists flatbeds by suitability and price. Color management software is essential to ensure that the digital representation is as accurate as possible. This can often be purchased with the scanner. RLG DigiNews December 1997 (Vol 3 number 3) has a technical review of color management software which is a good starting point. Transparencies can be scanned on a flatbed if it is capable of sufficient resolution and has a transparency adapter fitted that will shine light through the transparency into the scanning head. However, faster and potentially better results will be gained from a dedicated transparency scanner. These scan strips or mounted 35mm negative or positive transparencies to high resolutions. Scanning unmounted strips or single frame transparencies on a flatbed is difficult and time consuming because they have to be either placed in holders or taped to the scan bed to stop them moving in the heat of the light - using a transparency scanner can alleviate

some of this effort and would be a good investment if 35mm is a considerable part of the collection.

4.51 Digital cameras:

Digital cameras are developing for both the home and professional market and are priced from several hundred to thousands of pounds. 'Home use' cameras are aimed at non-professional users for taking general casual photography. There are two kinds of professional digital camera; the first has developed from medical and industrial uses and is a complete unit. The second is where the film from a traditional camera is replaced with computer sensors that transmit the image to a computer rather than to film; this is known as a digital scanning back. The first type has been around for longer and has been used in imaging projects for several years. Digital scanning backs are developing for professional photographers as a replacement for traditional film cameras. One of the advantages of the scanning backs is that they use the lenses and camera body of a traditional professional camera. Professional digital camera set-ups will generally require the operator to understand the basics of photography.

4.6 IN-HOUSE SCANNING UNIT:

The conversion of the materials can be done either in-house on specially purchased or existing equipment or sent to an external agency. Setting up a digitization unit gives the institution the value of equipment and trained staff for future projects and the movement and treatment of the materials can be closely controlled. Using an external supplier to do the scanning means that the equipment and expertise of a third party can be exploited while the project team concentrates on their specialist area of the project. Both approaches have their merits but there are certain situations where the choices are more clear cut.

Major reasons for sending materials to a external agency for digitization rather than attempting to scan them in-house include that the originals are not capable of being scanned successfully in-house (for example the equipment is excessively expensive) or

that the intended product is beyond the experience and abilities of the project - for example requiring advanced color management skills. As an example, the type of equipment used for the scanning of items such as bound books or microfilms tends to be so expensive that it may be difficult for a project to justify the expenditure on such equipment, particularly given the short life-span and high maintenance costs of scanning equipment. Other reasons for outsourcing may include where there is a large volume of work to be done in a short period of time or where the project has space, infrastructure or staffing constraints that preclude the setting up of in-house facilities.

Alternatively, the digitization manager may decide to use in-house resources for several reasons including that:

- The collection cannot be moved out of the institution.
- The collection is badly organized (organizing it well enough to send to an external supplier would be an excessive overhead).
- The digitization needs to be phased in small amounts over a long period.
- The digitization task is very simple.

There are some baseline infrastructure requirements for in-house digitization:

- A robust production level scanner which will be able to scan the originals to a suitable resolution.
- A powerful PC with lots of memory (at least 256Mb RAM) - or Mac equivalent.
- Plenty of system resources such as backup and write to media (e.g. CDROM) capacity.
- Software to assist the digitization.
- Experienced/competent staff to run the equipment and staff to oversee the process and/quality assurance.

This is assuming that the in-house operation wants to approach anywhere near the unit prices of production available from outside agencies. A further reason why many

digitization works are undertaken in-house is that the staff time, overheads and some consumables such as file storage can often be swallowed up by the institution and do not become apparent as a costed factor of the project, thus making this appear to be a cheaper option than out-outsourcing.

4.7 DIGITIZATION FORMATS:

Since computers have been employed for general business use the phrase 'Paperless Library' has been quoted as a goal for the modern Library and Information Centre. The same principal applies to the managers of archive drawings and photographic records, who have the facilities for the storage of their archives in digital format. The conversion of the original hard copy records into digital records is known as digitization. Broadly there are two approaches to achieve the required end product.

For line drawings the question needs to be asked as to whether the drawing is likely to be used and modified or is retained as an archive. If it belongs to an archival collection, the drawing is scanned and stored as a digital file on CD-ROM or DVD. If the drawing is to be used as a working drawing on a Computer Added Design (CAD) system, then the scanned data needs to be converted into vector data in an intelligent form i.e. in a way that replicates the same drawing if it had been produced originally on CAD.

Photographs are also scanned using the original negatives or transparencies if possible, if not then photographic prints. Old and damaged photographs can be cleaned and digitally repaired.

Three major issues related to storage of Digital information:

- The format of the digital data and any compression has a major effecting file size and a minor effect on quality.
- The resolution of the scanning is a balance between quality and file size.
- The storage medium can have an effect on retrieval times.

As computer storage is becoming less expensive, it supports to move towards higher standards, i.e. higher resolution. But it should be remembered that a relatively small increase in the resolution results in a proportionately high increase in file size. For example, a three color 8-bit image (RGB) at A4 is 100 Megabytes in uncompressed form at 600 dots per inch, but rises to 400 Megabytes at 1200 dpi.

Several raster-based file formats are available to store digitized information. Several new image file formats are also emerging. Many are limited to particular applications, such as digital capture devices or image manipulation programs. Others are destined for wider use, with their developers intending them to become official or defacto standards. It is a long and difficult process to create such standard formats and longer still for them to become widely used and supported.

Presently we have mainly six formats, which may be divided into two categories:

Open (Non proprietary) Formats : TIFF, PNG, GIF and JPEG 2000
Proprietary Formats : MrSID, DjVu, Genuine Fractals and PixelLive/VFZoom. Here I will discuss all of these formats along with their features. Some of the earlier common file formats are also discussed in brief.

4.71 Open (Non proprietary) Formats:

Any digitization project will need to consider the long-term usefulness and accessibility of the images and this means choosing a file that is both an established industry 'standard' as well as a non-proprietary format. Some of these formats are discussed bellow:

4.711 Graphic Interchange Format (GIF):

The Graphic Interchange Format is an 8-bit (and under) indexed file type. It only offers a range of 256 (or less) different colors. These can either be a standard selection or an image-dependent selection by user-choice. It was designed in the early days of the

Internet by Compuserve and works best for use with vector images using block colors, such as graphics, logos and banners. GIF uses LZW lossless compression which is a patented compression algorithm and for that reason should only really be used with caution. The amount of compression will depend totally on the type of image. A full color continuous tone image is unlikely to compress to less than 30% of its original size, however a solid color vector image should compress far more. The GIF file format supports layers allowing it to offer both transparency and animation.

4.712 Portable Network Graphics (PNG) file format:

The Portable Network Graphic (colloquially called 'PING') is an open raster image format. PNG is not so new. It was developed in 1995 as a replacement for the Graphics Interchange Format (GIF). It is normally used in either an 8-bit indexed version or as a 24-bit full color version, although there is also an infrequently used 48-bit version as well. It is a very versatile format, which offers either the advantages of lossless compression in full color or as a replacement of GIF in 8-bit form. It is supported by the W3C and IETF and expected to be released as ISO/IEC International Standard 15948. The latest version is PNG 1.2.

PNG has provision to support a number of different compressions, but only one is currently defined for use within the format, the lossless 'Deflate' compression (also found in zip and pkzip formats). Deflate uses a combination of LZ77 and Huffman encodings, both of which are patent free. PNG compresses better than GIF, saving 5-25% for equivalent (i.e. 8-bit) images, and in higher bit modes it achieves good savings over an equivalent uncompressed TIFF image. A small caveat, although PNG is defined as lossless, some applications that write PNG images do actually throw away a little image information in order to optimize the file size. This introduces loss and should be avoided. It can represent a lot more color information than the GIF. In addition to an 8-bit 'paletted' mode (256 colors, as in the GIF format), PNG adds two other modes, grayscale (up to 16-bits) and true color (up to 48-bits). PNG also allows color profile information to be stored within the file, enabling applications with color management systems to

accurately reproduce the color on the screen or in print. Besides these PNG is Gamma correction, the PNG can record the gamma of the image (i.e. the brightness level), as it was set when the image was created. This feature enables the image to be automatically adjusted to display well on different monitors. Alpha Transparency Channel is also a prominent feature of PNG. GIF enables one of its 256 paletted colors to be declared transparent. At the same time PNG goes even further. In addition to its palette transparency, it offers a full 'alpha channel' ranging from no transparency (i.e. opaque) through 254 levels of partial transparency to full transparency). This feature is primarily used in graphic design, supporting image fade-outs, drop-shadows, and the seamless overlaying of images. The next one is two-dimensional interlacing. The GIF's interlacing feature offers a way of progressively displaying (streaming) the image. It operates in one dimension, rather like a Venetian blind. PNG offers optional interlacing in two dimensions, building up the image horizontally and vertically at the same time, in 7 distinct passes (formally known as 'Adam 7'). It also interpolates the space in between while waiting for the actual data to arrive. Interlacing a PNG slightly slows its delivery, but enables the image to be understood by the viewer before it is fully downloaded.

4.713 Tagged Image File Format (TIFF):

By the mid-nineties there was some discussion about replacing CompuServe's GIF format, but the immediate prompt for PNG's development was a patent dispute. Unisys asserted their rights to the LZW compression that lay at the heart of the GIF format, forcing those developing software to pay royalties whenever they made use of GIF/LZW. In response, the new PNG format was hastily drafted. It avoided using LZW and other proprietary technologies and took the opportunity to improve on the GIF's functionality. It extended the bit-depth (from 8-bit to 48-bit), offered better color and gamma support, and better interlacing and transparency features.

In every respect the PNG exceeds the GIF. But one feature it doesn't support is animation. This is because there is a complimentary MNG (Multiple-image Network Graphics) format, finalized in 1999 and beginning to find some support.

In practice PNG was called 'the new GIF', since it seeks to replace it, but the PNG goes far beyond the GIF. Although the GIF is technically lossless in its compression, it is inherently lossy since it first shoehorns an image's color information into a palette with a maximum of 256 colors. The PNG might just as well be regarded as the new TIFF, since it offers lossless storage with an equivalent bit-depth. Despite its head-start and growing usage, there was some possibility that the PNG might not fully take hold if the more flexible JPEG 2000 achieves a good take-up and, ironically, the GIF turns into an open format.

It is still not widely used and it has taken some time for Web browsers and image application software to support it. Now PNG files have reasonable support among the leading browsers and can be created and manipulated within many image applications.

4.714 Joint Photographic Experts Group File Interchange Format (JPEG or JFIF):

The common JPEG compression and its corresponding file format were developed in the late 1980s by independent members of the Joint Photographic Experts Group (JPEG). In its core or 'baseline' form, the common JPEG is lossy. A later, lossless, version was developed (the JPEG-LS) but is rarely used. More successful was the addition of a progressive (streamed) display for the JPEG. This formed part of the original standard, but not widely implemented until 1996. It divides the file into a series of scans of increasing quality, enabling the image to build up progressively as it is displayed.

JPEG is not actually a file type, but a type of compression proposed by the Joint Photographic Experts Group. It is a lossy compression and provide the best quality and lowest file size for continuous tone images. The amount of compression given to the file is chosen at the time of saving the file and allows for variation in quality against file size, as a rule of thumb, it is normally considered that a file compressed with JPEG to 10% of its original size will be visually acceptable with no obvious compression artifacts.

However it is common if required, to compress right down to 2-4% if the lower quality is acceptable.

It is an open raster image format described by the ISO/IEC standard 15444, and ITU standard T.800. It is intended as a replacement for formats using the JPEG compression, particularly the JPEG/JFIF format commonly used on the Web, It is used within the JFIF file format that uses the file extension .jpg and we colloquially call the 'JPEG'. The baseline version of JPEG 2000 is known as 'JPEG 2000 Part 1', and is usually given the extension.jp2 or, less often, .j2k.

From the late 1990s work began on a successor standard, the JPEG 2000 - which defines both the compression and its corresponding file format. JPEG 2000 uses state-of-the-art wavelet compression techniques that are capable of both lossy and lossless compression. Although JPEG 2000 includes some patented technology, efforts have been made to keep the baseline version license and royalty free. As a consequence, some other proprietary compressions and display features have been reserved for later versions of the standard.

JPEG 2000 Part 1 became an official standard at the end of 2000 and an extended version .jpx or 'JPEG 2000 Part 2', was approved in the following year. A number of other parts of the standard are being developed. These introduce new features (some of which are patented); deal with other types of images (e.g. the 'Motion JPEG' in Part 3) or with new applications (a wireless JPEG' in Part 11).

Some of the features available within the baseline standard (Part 1) are:

It supports very good compression, both lossless and lossy. The common JPEG compression was always lossy. In contrast, JPEG 2000 supports both lossy and lossless compression. JPEG 2000's compressions are based on wavelet techniques. The wavelet compression turns the image into waves and then generates a series of increasingly

simplified versions. A lossless JPEG 2000 will contain all the information necessary to rebuild a complete wave, while a lossy JPEG 2000 image will make do with the simplified versions. JPEG 2000's compression is more efficient than other common compressions. It will deliver lossy images 3-5 times smaller than comparable JPEG images. Lossless JPEG 2000 images are necessarily larger, but are still generally half the size of the original uncompressed raster image. This is better than the lossless LZW compression (used in GIF and optionally in TIFF) or Deflate (used in PNG).

The common JPEG divides an image into very small (8x8 pixel) blocks and processes these independently. Where the compression is high (and quality low) the boundaries of the blocks begin to show. In contrast, the JPEG 2000's wavelet compression processes much larger areas of the image at once - sometimes the whole image. This avoids any blocking. The wavelet compression is also able to make a distinction between significant detail in the image, like edges, and less significant areas, for example, where there are slight variations in the color. At very high compression JPEG 2000 will still introduce artifacts (visible distortion in the image), but by concentrating its compression on the less significant parts it gives much better overall quality than a JPEG of the same file size.

4.714 Comparison between TIFF, JPEG and JPEG 2000:

Detail of original TIFF image, full image = 776K

Detail of JPEG compressed image, full image = 23K

Detail of JPEG 2000 compressed image, full image = 23K

4.72 Proprietary Formats:

While PNG and JPEG are the new general open formats that will be seen in increasing numbers on the Web, there are other new formats that are worth being aware of. These are proprietary and tend to be used for specialized tasks. Four of these in some depth below are: MrSID, DjVu, Genuine FractalsTM and PixelLiveTM (formerly VFZoom). The MrSID and DjVu formats specialize in encoding particular sorts of images.

MrSID deals with large pictures, plans or maps, DjVu, with documents containing a mix of text and image. Genuine Fractals and PixelLive specialized in a particular task, as scaling or enlarging images. If open formats can suffer through lack of use or patchy or inconsistent support, proprietary formats can be vulnerable to commercial pressures, since they are tied to the fortunes of one company. More positively, they can also benefit from significant investment in their development, a competitive environment that rewards innovation, and from synergies with other, associated commercial 'products'.

The formats below have something useful, but their proprietary nature should urge some caution. MrSID, DjVu, and PixelLive are worth considering as Web display

formats (though each currently requires a special viewer). However, they may not be the best choice for image archiving.

4.721 MrSID:

Multi-resolution Seamless Image Database (MrSID) file format shows the file extension '.sid'. It is designed to compress huge images seamlessly and allow selective delivery and decompression.

Originally proposed as a format suitable for many different purposes, MrSID has concentrated on Geospatial applications, to which it is well suited. It is supported by the leading GIS software and used by official mapping agencies like the United States Geological Survey (USGS) and the National Imagery and Mapping Agency (NIMA), primarily as a delivery format. Long described as 'visually lossless' (i.e. lossy, but without too much obvious degradation), the latest version (3.0) introduces a truly lossless MrSID.

Like JPEG 2000 and DjVu, MrSID is based on wavelet compression. It is a good advertisement for this form of compression, achieving efficient and high quality lossy compression which varies according to the image content and color depth, but averages 20:1 for 8-bit grayscale and 50:1 for 24-bit color. Of all of the wavelet-based formats, it best illustrates the zooming potential offered by the wavelet's multi-resolution approach. Instead of storing a handful of predetermined resolutions, as some other proprietary formats do, MrSID includes all the information necessary to rebuild the image at any size (i.e. resolution) up to 100%. The process of rebuilding the image can be observed when you use the MrSID viewer's sliding zoom facility: the image is instantaneously resized, but it takes a moment or two for all the fine detail to fill in.

Although it is a proprietary format, once displayed and zoomed a MrSID image can be easily 'exported' (resaved) into a TIFF format at any chosen resolution. Its developers promote it as a storage format, because of its efficient compression, but it is believed that its chief advantage is its delivery and display potential.

Several big digital library projects have chosen MrSID to deliver large format cartographic materials, including the US Library of Congress.

4.722 DjVu:

DjVu is a screen/Web format and is more suited to 'mixed documents' (i.e. text and image) than to individual images. It was developed within the AT&T research labs in the mid-nineties, and then acquired and commercialized by LizardTech in 2000.

Although it is a proprietary format, DjVu's specification is available for non-commercial use and there are several open source implementations of the decoder (viewer) available. Open encoders have also been developed, but these cannot really compete with LizardTech's software, which has kept the best compression engine to itself.

DjVu produces files that are 10-30 times smaller than a comparable GIF or JPEG, between 500 and 1000 times smaller than TIFF, and 50 times smaller than PDF, although results will vary considerably depending on the nature of the document and whether it is in color or black and white.

The reason DjVu is so good at compressing documents is that it divides each page image up into different components, treating hard lines and continuous tones in different ways. The lines on the page (e.g. text or drawn lines) are identified and pulled into a separate layer. They are then compressed using a bi-tonal (black/white) compression technique. The softer tones and colors are pulled into other layers and subjected to different compression techniques.

DjVu's bi-tonal layer is called 'DjVu Text'. It is kept at a high resolution (300dpi) and compressed and encoded using a version of the official JBIG2 standard (ISO/IEC

14492), which is more efficient than the Group 4 compression used by TIFF or PDF for bi-tonal compression. DjVu's bi-tonal encoding can be lossless.

The tone and color layers are called 'DjVu Photo'. There are usually two, a foreground and a background layer. These have lower resolutions (100dpi and 25dpi respectively) and are compressed using a wavelet transform very similar to the one used in JPEG 2000 or MrSID.

The 'DjVu Text' and 'DjVu Photo' layers can be saved and used independently, but are commonly kept together within the 'DjVu Layered' format - usually just referred to as the DjVu format. This layered DjVu format is inherently lossy.

DjVu handles multiple pages in one of two ways. It either bundles everything into one file, like a PDF, or it stores each page as a separate file. Metadata can be written into the file, and OCR'd (optical character recognized) text can also be included to facilitate text highlighting or searching.

This format can handle fairly large sized documents (up to 32,000 x 32,000 pixels equivalent to 100 x 100 inches at 300dpi), so is suitable for some large drawings, maps or plans. Anything bigger or purely photographic is better handled by MrSID or another format. DjVu does a particularly good job of representing handwritten letters, manuscripts and early printed materials. It provides crisp, clear text or line art while preserving the look and feel of the material on which it was written or printed.

4.723 Genuine Fractals and PixelLive/VFZoom:

It is useful to consider LizardTech's Genuine FractalsTM and Celartem's PixelLiveTM/VFZoom together, since they perform a similar task – enlarging images. However, they do so in slightly different ways and using different technologies. Strictly speaking, they are not raster formats, but we include them here because they are used to

encode types of images typically stored as rasters (i.e. continuous tone images, like photographs).

Most of the formats discussed above enable zooming, but they are only really intended to be zoomed up to full size (100%, pixel for pixel). Anything beyond this and the images will become pixelated (i.e. show their pixels). With a raster format, the only way to avoid this is to interpolate - to insert new pixels in between the existing pixels. Imaging applications like Adobe Photoshop and JASC Paint Shop offer good interpolations, but there are also dedicated interpolation products like S-Spline (http://www.s-spline.com/) and Xfile.

Genuine Fractals and PixelLive/VFZoom offer alternatives to interpolation-converting the raster into a different form of information (fractal or vector) before enlarging and re-rasterising it.

Instead of leaving the image as a set of pixels, Genuine Fractals breaks the mage into small shapes (fractals), which are described mathematically and can be redrawn at a larger scale. This is done by opening up a raster image within the Genuine Fractals software or a Photoshop plug-in and specifying its new dimensions. Depending on the size and complexity of the image, the fractal encoding and enlarging can take a long time, often a number of minutes. Once enlarged, the images can be left in the Genuine Fractal "STiNg" format (with an .stn file extension) or can be re-saved into another format for printing or Web display.

In addition to resizing, a fractal approach has the potential to offer efficient image compression, by finding identical or near-identical shapes within the image and replacing them with the same equation. This functionality is available within Genuine Fractals, but the format is being promoted more for its scalability rather than its compression. It has attracted a strong following within the professional photographic community where it often used to enlarge images for printing.

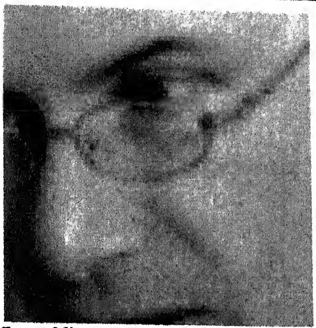
Celartem's new PixelLive format (launched mid 2003) is based on its earlier VFZoom format and still retains a .vfz or .pfz file extension. Like Genuine Fractals, it also avoids the simple raster matrix. In PixelLive the pixels are converted into vector information (e.g. lines and shapes and fills), which, like the fractal, are capable of being mathematically described and drawn at a bigger size. The vector nature of this format is most easily seen when low quality versions of PixelLive images are enlarged.

Unlike Genuine Fractals, the PixelLive format serves as both a scaling and a display format (using a free PixelLive Viewer). What is particularly interesting about PixelLive as a display format is that the raster image is encoded at its original dimensions and delivered to the user at this size. Any vector enlargement takes place within the user's viewer, where it looks and acts like a zoom function. The earlier VFZoom format even allowed the user to resave and print the image at its increased dimensions. However, the newer PixelLive format removes this functionality, making a greater distinction between the activity of scaling or resizing the image (which can now only be done within a PhotoShop plug-in called pxl SmartScale) and merely zooming the image when viewing (which is what happens within the PixelLive Viewer).

The PixelLive format has other features that are intended to support its delivery and display. It offers a quality setting with 6 discreet levels (0-5). At level 5, the image is losslessly encoded and capable of being returned to its original raster format. But it is also possible to encode and view the image without its higher levels. This will reduce the file size, but will obviously involve some image loss. Another feature is PixelLive's password protection option. When saved it is possible to set a username and password, which the user must enter if they are to view the image. A password-protected PixelLive file is given a .pfz file extension.

4.7231 Comparison between Genuine Fractals and PixelLive/VFZoom:

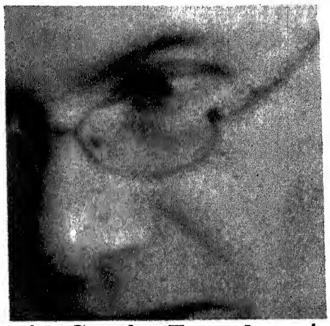
The Comparison between Genuine Fractals and PixelLive/VFZoom is shown with the halp of a captured photograph. The difference is clearing indicating the difference between the two and with the actual magnified form.



Magnified 800% (small sample)

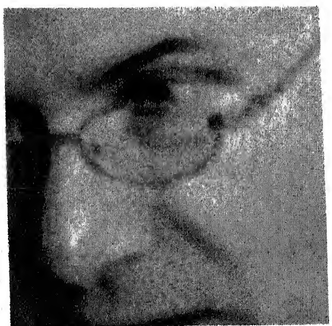
Interpolated in Photoshop (without sharpening) to give 4000 x 4000 pixels (small sample)

Rescanned to give 4000 x 4000 pixels (small sample)



Scaled up with Genuine Fractals to give 4000 x 4000 pixels (small sample)

Scaled up with PixelLive (without SmartScale enhancement) to give 4000 x 4000 pixels (small sample)



Scaled up with PixelLive (with default Smart Scale enhancement) to give 4000 x 4000 pixels (small sample)

4.73 File Format for Specific Purposes:

There have been a number of image file formats that have been used. Of course, every year, this choice gets larger and larger as new file formats are introduced and it is not always immediately clear which file format is the best one to use in any particular case.

The choice should depend on a number of factors, which may vary according to how the users intend to use the file and what will be the pattern of usage in the information center. Each stage of the process, from capturing until delivery has its own requirements that may affect this choice. A brief look at some of these factors and guidelines for making the best choice from what is available.

4.731 File Formats for Capture:

This is the first step in the digitization process. When capturing images, it is important that they are all created at the highest possible quality and at a size appropriate for all subsequent uses. Errors at this point will certainly compromise the quality of the whole project and the only recovery option will be to go back and re-capture the original.

All digital capture devices originally capture values of Red, Green and Blue. The number of different describable colors (or tones of gray) will depend upon the 'bit-depth' of the device. Any modern device will be able to capture in at least 24-bit color (or 8-bit B&W), although many modern devices now offer to capture at higher bit depths, right up to 48-bit. The suggested format for this purpose is: TIFF or the proprietary format of capture device.

4.732 File Formats for Master Archive:

The requirements of a file format for archiving are the same as for creation, the suggested formats: TIFF and PNG.

4.733 File Formats for Optimization and Manipulation:

All image optimization and manipulation is undertaken within image processing software. Whilst carrying out this work, it can be useful to save the image in the

proprietary format of the image processing software. Suggested formats for this purpose are image processing proprietary formats such as PSD for Photoshop, PSP for Paint Shop Pro and PNG for Fireworks. However TIFF is still a good choice if the increased functionality of the proprietary formats is not required.

4.734 File Formats for Delivery:

Choosing the correct image file format for delivery probably poses the hardest decision with the biggest variety of choice. These are just some of the issues that will need to be considered:

- What is the intended use of the image after delivery?
- How much image resolution is needed to convey the intellectual content to the user?
- On what output device is the image going to be used - monitor, printer, projector?
- What are the capabilities of the output device? What bit depth can it handle? What is the required resolution?
- What bandwidth is available for delivery?
- Is the image for photo-realistic or presentation use?
- How is the image going to be delivered? CD-ROM, tape, WAP, Internet (modem or LAN/WAN connection)?
- Is there a requirement to add any watermarking or deal with any other digital rights management issue?
- Do the users require the image to be provided with any color profile or other color management information?
- With so many considerations, combined with the proliferation of file formats, each designed for a specific use.

4.735 File Formats for Web Delivery:

For most digitization projects, the most common delivery format is simply a monitor with the images viewed through a Web browser interface. This makes the choice of file format easy as the current selection of Web browsers only support a small range of image file formats (JPEG, GIF & PNG), although this range can be extended with the use of the appropriate plug-in.

Delivering images through a Web browser has some inherent advantages and unfortunately some challenges. The main advantage is that (in common with all monitor delivery) images naturally look 'good' on a monitor where their perceived 'brightness' (the light is being transmitted to you, rather than reflected) hides many small deficiencies in quality that would compromise quality if the image was printed. On the other hand, present browsers have only limited image-viewing capabilities and are unable to 'zoom' in and out of the images. This means that delivery is limited to images with pixel dimensions that fit within the user's browser - suggested standards at present are to design Web pages to a size of 800 x 600 pixels giving standard image sizes of approx 512 pixels on the longest edge.

The biggest limitation on the quality of images delivered on the Web and the main influence on 'choice' is the need for them to be compressed to a size that makes their delivery over the limited available bandwidth possible. All the file formats supported by Web browsers provide compression, however the amount and method of compression varies.

4.736 Web browsers currently support the following file formats:

JPEG (JFIF) - JPEG is not actually a file type, but a type of compression proposed by the Joint Photographic Experts Group. It is used within the JFIF file format that uses the file extension .jpg and we colloquially call the 'JPEG'. It is a lossy compression and will provide the best quality and lowest file size for continuous tone images. The amount of compression given to the file is chosen at the time of saving the file and allows for

variation in quality against file size: as a rule of thumb, it is normally considered that a file compressed with JPEG to 10% of its original size will be visually acceptable with no obvious compression artifacts. However it is common if required, to compress right down to 2-4% if the lower quality is acceptable.

GIF: The Graphic Interchange Format is an 8-bit (and under) indexed file type only offering a range of 256 (or less) different colors (these can either be a standard selection or a image-dependent selection by user-choice). It was designed in the early days of the Internet by Compuserve and works best for use with vector images using block colors, such as graphics, logos and banners. GIF uses LZW lossless compression which is a patented compression algorithm and for that reason should only really be used with caution. The amount of compression will depend totally on the type of image. A full color continuous tone image is unlikely to compress to less than 30% of its original size, however a solid color vector image should compress far more. The GIF file format supports layers allowing it to offer both transparency and animation.

PNG: The Portable Network Graphic (colloquially called 'PING') file is an open source 'standard' that was introduced to overcome the possible patent problems associated with the GIF format. It is normally used in either an 8-bit indexed version or as a 24-bit full color version, although there is also an infrequently used 48-bit version as well. This makes it a very versatile format offering either the advantages of lossless compression in full color (as an archive format) or as a GIF replacement in 8-bit form. However it cannot compete with the JPEG in terms of producing high quality and small, full color images for viewing on the Web. The compression available from PNG in 24-bit mode is typical for a lossless compression providing a file of about 60-75% of the original size and in 8-bit mode it is much the same as GIF. PNG supports transparency (even variable opacity, although browsers do not!) but is not able to provide animation. Suggested formats and relevant uses are JPEG, PNG, and GIF File Formats for PowerPoint or Other Multimedia Programs.

The main influence on choice will be the available bandwidth for the delivery of this material. If there are still some bandwidth or performance restrictions (Internet, need to save on floppy or deliver on slow machine) then it will make sense to use some of the file formats suggested for Web delivery, however if the presentation is to be delivered locally from a fast machine, then there is no reason to not use images of a correspondingly higher quality. Suggested formats for monitor delivery are JPEG, PNG and GIF.

4.8 COPYRIGHT ISSUES:

There are many interdependent and interacting factors to be weighed in selecting materials to digitize. The specific choices that result from the selection process will reflect subjective judgments, any of which may change over time. Nuance assessments, ambiguity, and shades of gray are all to be expected.

Questions concerning copyright, however, are far more clear-cut. Simply stated, if a proposed digitization work involves materials in the public domain, the work can proceed. If the source materials are protected by copyright but rights are held by the institution or appropriate permissions can be secured, the work can move ahead. If permissions are not forthcoming for copyrighted sources, however, the materials cannot be reproduced. Copyright assessments thus play a defining role with regard to digitization projects.

Copyright issues in the digital environment are still very much in flux and have provoked ongoing international discussion. While the broad thrust of digital technology is toward enhanced access, diminished costs, and more versatile capabilities, it is far less clear that copyright law will likewise encourage wider use. The legal strictures applicable to a particular project will vary depending on the country in which the project is based, the country in which the source materials were produced, and prevailing international agreements. Different kinds of materials, moreover, usually pose different types of rights-management issues. The performance rights associated with musical scores, for example,

or exhibition rights for films, differ from rights for nonperformance materials such as electronic journals or documentary photographs. To complicate matters, all these rights are susceptible to change over time.

Digital projects must be undertaken with a full understanding of ownership rights, which is difficult as they often are to ascertain, and with full recognition that permissions are essential to convert materials that are not in the public domain. Rights that must be negotiated with the copyright holder often entail fees. The institution hosting a project may also have policies and procedures that inform intellectual property negotiations. The legal office of most institutions can provide guidance. The Internet site IFLA: Copyright and Intellectual Property Resources is a good resource for maintaining current awareness. It includes articles, reports and white papers, discussions, and information about organizations related to copyright issues, intellectual property in general, and electronic distribution of intellectual property.

4.91 PROCESS OF DIGITIZATION:

Now it is clear that digitization is an exciting preservation option while providing unparalleled access available to all. The technology is advancing rapidly and this raises the question of accepted standards for digital preservation technology. The standards for image capture, resolution, data transfer protocols, indexing, access, and file types must be of sincere consideration. Some of the important considerations are as follows:

4.911 Defining your requirements:

The first consideration before even thinking about equipment is to define the end use of the images. On this will depend what resolution of images and what file sizes are required, and hence the equipment and time scales involved. It is also essential to consider how you will store and access the scanned images and to ensure that your budget includes the cost of the right software to meet your requirements.

4.912 Choose the right equipment:

It is important to remember that scanning technology advances quickly and enhancements are regularly incorporated into new products meaning hardware has a high depreciation rate and may quickly become obsolete and uneconomical to maintain.

The main types of capture devices available are:

3.9121 High end A3 flatbed scanners:

These are suitable for all types of photographs, transparencies, negatives and pages up to A3 size that may be laid absolutely flat. They are best used to produce files of 20 Mbyte plus. They are not suitable for bound volumes, glass plates, mounted slides, formats larger than A3 or, because they use very bright light, anything that is in danger of fading.

3.9122 Drum Scanners:

These are used by reprographic houses. Whilst they produce very high quality results they are expensive and the originals have to be fastened around a drum, which means they need to be very flexible and unmounted.

3.9123 Medium format scanning backs:

Scanning backs are essentially the devices that convert a medium format conventional camera into a hi-resolution camera. With the right accessories, they are ideal for capturing items that cannot be placed on a scanner.

3.9124 Digital cameras:

Digital cameras come in a variety of standards. To be suitable for digitization work these must be of a professional standard and capable of 18 Megabytes plus, with interchangeable lenses and accessories.

3.9125 5 mm scanners:

These would seem to be ideal for collections made up of slides only. However, many of them are aimed at the domestic market and will not be robust enough for any

reasonable sized collection. They often struggle to produce up to 18 Megabyte files of a good dpi.

3.9126 Workstations:

All capture devices will need a dedicated workstation of good specification and with monitors capable of proper calibration, appropriate scanning software, a CD writer and the ability to store a reasonable number of images, either on the hard drive or server.

3.92 SET UP YOUR SCANNING LABORATORY:

Before digitization can start there are a number of things needed to consider. Over and above the environment needed to run computers the following are essential:

- A stable power supply - fluctuations can ruin scans and make consistency impossible.
- The scanning equipment must be installed in a proper clean area - dust and dirt is exaggerated when viewed on-screen.
- The workstation must be co-located with the scanner - The person who digitize, must be able to view the scan on–screen while working.
- Check links to networks or servers are available - you will be moving large files.
- If cameras are used the ambient light should be fully controllable - overhead lighting or the variability of daylight will add a permanent color cast and make consistency impossible.
- The studio should have solid floor to prevent vibration - no digitizing is instantaneous.
- There must be sufficient room and suitable surfaces to clean items, handle and store them before, during and after scanning.
- The area must be secure.

3.93 DECISIONS ON PROCEDURES:

3.931 Sorting and Cataloguing:

Every item to be digitized must be clearly identifiable. If the digital file name is not marked or labeled on the original item, then it should carry some other unique identifier, such as an accession number, which can be associated with the file name during the scanning process.

3.932 Job Specification:

Before starting the work, the rules and conventions must be decided for the job. For example:

- What file type and final size or resolution is required?
- Should monochrome prints be captured as RGB or greyscale.
- Should images to be rotated for upright orientation?
- Should operators crop to edge of item or inside edge of print/negative? It should be specified whether captions are to be included if they are located outside the regular cropping convention, or exclude all or some types of captions.
- Should color casts to be corrected? Film transparencies and negatives very often have a color cast created by the film material, lighting conditions or processing - or a combination of all three. There are methods for detecting the presence of and correcting such irregularities.
- Is re-touching to be done? If so, to what extent, at what screen size and what is the maximum time to be spent on each item?
- Is the original unprocessed scan to be retained?
- If visible or digital watermarks are to be applied - to which resolution versions and when in the process?

3.933 Prepare your materials:

Before scanning you need to prepare your equipment and originals.

3.9331 Batching:

Where possible group original items in batches of same size/format materials, say 100 - 150 in a batch. This will make the work easier to manage and improve productivity. It also means quality issues can be more easily isolated and corrected.

4.9332 ICC Profile Checking:

Ensure the monitors are all calibrated and profiled to the same standard, that the capture devices are individually profiled and that the final files are all tagged with the required device independent ICC color profile.

4.9333 Pre-cleaning:

If materials have been in storage for any length of time a degree of controlled cleaning may be desirable. Cleaning processes should be carried out well away from the digitization area using lint free cotton and surgical gloves.

4.9334 Oil mounting:

Oil mounting is a specialist technique that can significantly reduce effect of scratches, ground in dirt and other damage to negatives or transparencies.

4.9335 Adjust and check your scans:

This is an important step and often overlooked in the planning. Once the image is captured it must have basic checks carried out, either immediately after capture or in batches at a later time.

4.9336 Color range and balance:

Check that the capture device's sample range has been utilized and neither end of the tonal. If color cast correction is to be carried out, this is the time to do it.

4.9337 Cropping and rotating:

Carry out cropping as decided in your job specification and, if necessary, rotate the image for correct orientation.

4.9338 Re-touching:

Carry out re-touching to the standard decided in the job specification - this needs training and/or experience. Retouching on photographic image will always show if not done to a professional standard.

4.93391 Compression:

Where required create a compressed version of the image.

4.93392 Multi-pack creation:

Where required create screen resolution multi-pack versions.

4.93393 Water-marking:

Where required apply visible or digital watermarking to specified versions of the image.

4.93394 Quality inspection:

It is advisable for a quality inspection of say 20% of output to be carried out by a qualified person who hasn't been involved in either scanning or post-processing.

4.934 Ideal Steps:

Ideally a digitization process should involve following steps:

1. Image preparation;
2. Preparation of the volumes, issues and pages of the document for scanning;
3. Scanning pages of the document;
4. Editing images of the document;

5. Using Optical Character Recognition Software (OCR) on images to edit text versions. (OCR is the process whereby a computer program "reads" the text from an image of a document and converts it into ASCII text);
6. Creating a searchable database of the text; linking text to images;
7. Mounting on Web.

4.9341 Image preparation:

The first step is to take all measures that are necessary to preserve the hard copy of the documents. A physical review of the document revealed much variance in the condition of the pages and that, while some could withstand scanning without any preservation, other pages would not. Consultation took place with the preservation library assistant in the rare book section at the university library and procedures were set up to follow:

1. Trim each page so that no rough edges are left;
2. Mend rips and tears with acid-free binding tape;
3. Put each issue in an acid-free envelope and store each volume in an acid-free storage box.

4.9342 Important factors to consider for format:

At this stage it is necessary to decide the most appropriate file format for the online version of the document. The integrity of each page had to be maintained and in order to do this, save the pages after scanning as images. There are two primary options:

1. Save the image as a PDF (Portable Document Format) file using Adobe Acrobat; or
2. Save the image as a GIF (Graphical Interchange Format) or JPEG (Joint Photographic Experts Group) file.

It is investigated to save the images as JPEG files, which would be most appropriate. JPEG will allow the entire project to be uploaded as an HTML file and furthermore, JPEG retains more information as it compresses an image:

- Each page had to be kept intact so that the user could view the page online just as it appeared in the paper copy;
- User-friendliness was important. Therefore, the online version had to be searchable, had to have a minimum of side-to-side scrolling, had to load fairly quickly and the pages had to be clear and easy to read;
- Optical Character Recognition (OCR) would work on scanned images of the document. This was a consideration as several different fonts were very small and that cannot always be recognized by OCR programs;
- With standardization the maximum number of users should be able to view it.

4.9343 Scan pages of the document:

Having determined the type of file to be used for the online version, several preliminary scans were conducted to determine the appropriate image type, brightness, contrast, height and width, as well as the length of time it would take for each page to be scanned. To ensure uniformity in scanning the following parameters have been investigated for set up:

Height: 12.65 in.

Width: 8.18 in.

Brightness: 125 (approx.)

Contrast: 130 (approx.)

Image Type: Black and White Photo

File Type: JPEG

Image Quality: 600 dpi

The process of deciding what to digitize anticipates all the major stages of project implementation. Digital resources not only depend on the nature and importance of the original source materials but also on the nature and quality of the digitization process itself. Besides this how well relevant information is captured from the original text and is organized, indexed, delivered to users, and maintained over time.

For taking a decision on to the selection of documents, these are presented in a sequence that moves from relatively abstract assessments of intellectual value to nuts-and-bolts issues concerning whether available resources and technology can provide a product that meets expectations. In practice, the pieces interact in ways that are often complex. Decisions about what to digitize must first and foremost address the intellectual value of the original sources. We are likely to be able to convert only a small percentage of existing scholarly materials to electronic form, and doing even this will require substantial investments. We therefore need to determine what it is truly worthwhile to convert.

Some scholarly resources are heavily used; others are consulted infrequently. With only limited funds available for reformatting, types and levels of use can help to shape priorities.

A person reading a book, looking at a photograph, or consulting a manuscript encounters few barriers to use. One might have to handle an object carefully, or use a magnifying glass to read fine print, but in general the work is immediately approachable. The same resource, when digitized, should be equally accessible and approachable. Ideally, the electronic version will also permit new kinds of use and more sophisticated types of analysis.

Decisions to digitize must take into account the physical size, nature, and condition of source materials as they affect the characteristics of the desired product. They must likewise address whether available means of conversion can satisfy expectations for the result. Projects must also, from the very first, consider how users will be guided through the electronic version.

After an extensive analysis of the above mentioned issues it is sorted out that a librarian or an information officer should consider following points while planning for the digitization of documents in his library and information centre:

Assessment of scholarly value of Library material:

Materials with marginal scholarly value are best left in their original form or made accessible in a less costly manner. Scholarly value, of course, is a subjective assessment and even the most marginal materials can support some kinds of research. Most users, nonetheless, would opt for electronic access to original monographs rather than to derivative works, or to the papers of a prominent scholar over the administrative records of a university department. Bibliographers regularly make purchase decisions that reflect their evaluation of the intellectual quality of single items or collections of materials. Similar judgments apply in choosing what to digitize.

Digitization and the intellectual value of the digital material:

Scholarship can be facilitated when texts are made fully searchable by rekeying (retyping) them or by employing OCR software. Comparisons between successive drafts of a text and the final published work, for example, or with later editions and translations, are vastly simplified when the words and phrases are searchable. A concordance or thesaurus is likewise most easily mined when it is in searchable form. Electronic texts can be moved readily from one environment to another (from the World Wide Web onto the hard drive of a personal computer, and then into a word processing program, for example), shared with other users, and manipulated and reconfigured for multiple purposes. Digitized prints, drawings, and other visual resources can be viewed in groups at low resolution or inspected individually for very fine detail. Digital charts and tables, appropriately coded, can be loaded directly into statistical software packages for additional analysis.

Electronic accessibility of body of information of original books, manuscripts, photographs, or paintings:

A collection of thousands of portrait images, however promising a resource, might be nearly unapproachable because of its size and the condition and dimensions of individual items. Well-indexed and in digitized form, however, the collection could

be searched with relative ease for images of a particular person or for some indexed characteristic (the country from which the portrait originates, for example). Likewise, the digitization of large-format architectural drawings could enable comparisons of small- and large-scale drawings, different views of the same architectural feature, or sequential phases of construction.

Increased value of the combination or aggregation of original sources increase their value:

Digitizing related scholarly monographs, like building a coherent collection of paper copies, can strengthen the context within which each title is approached. Ephemera-leaflets from a political campaign, for example-are often most useful when studied in the aggregate, as are posters, broadsides, and popular literature. Harvard has digitized daguerreotypes from thirteen repositories to facilitate the combinations and comparisons that are otherwise precluded by the fragility, value, and dispersion of the original images.

Popularity of digitized source material which is being consulted among scholars:

Intensive use does not automatically make a collection a good candidate for digitizing. If the primary audience is local, for example, and if competition for a particular resource is not a problem, access may already be sufficient. Ephemera produced by a community political organization may be of great interest to local scholars and of limited value to a worldwide audience. On the other hand, if use is heavy and widespread, digitizing may at once guarantee convenient and reliable access, and make it possible for some institutions to discard their original copies. The JSTOR project, through which a large array of core scholarly journals is being made accessible in digital form, is a prime example of an initiative focusing on high-use materials.

Digitization may provide wider access:

Low use may signal that a collection has marginal intellectual value, but there are many other reasons for valuable materials to have generated little interest. A

collection may be held in a remote location, for example, or be owned by an institution with highly restrictive access policies. Bibliographic records may be poor, as is often the case with pamphlets. The value of digitizing such materials may go beyond the simple fact that the resulting files can be widely distributed. Broader access, as it creates a new community of users, can also facilitate more active scholarship.

Physical condition of the original materials limit their use:

Some resources are too fragile to be consulted. Aging newspapers or palm-leaf manuscripts that break at the slightest flex simply cannot be browsed. In such cases, a digital copy might be provided to improve access, and a microfilm or other photographic surrogate made to ensure long-term survival. (Film can be made from a digital file or vice versa.)

Sources may also be at risk because of high user demand or extraordinary monetary value:

A nation's founding documents, glass-plate negatives of vanished architectural sites, or rare maps may benefit from the creation of digital copies that satisfy the purposes of most users. These files do not necessarily need to meet archival standards. They are created to protect the originals from handling.

Related materials are so widely dispersed that they cannot be studied in context:

Cooperative efforts to digitize disparate pieces of a greater whole can create or restore a more usable collection. Papyrus fragments, a prominent individual's far-flung correspondence, scattered photographs of a particular subject or by a specific photographer and broken serial runs are among the many materials whose coherence, accessibility, and scholarly utility can be enhanced through digitization.

Manageable size and format of Digital files:

Digital resources need to match users' technical capabilities and equipment. Most require Internet access and standard web browsers, or a CD-ROM drive. Images delivered to the Internet in formats other than JPEG or GIF require additional software for viewing or printing. Even when electronic resources are optimized for on-screen delivery, some network connections, particularly those via modem, are still far too slow to support browsing of digital collections at satisfactory speeds. And scholars in some locations may lack training opportunities or the ongoing technical support needed to take advantage of the electronic environment. These limitations, however, are not necessarily reasons to rule out digitizing. The worldwide trend is toward greater capabilities. Moreover, the more important the resources available electronically, the greater the incentive to acquire the network, viewing, and printing technology necessary to use those resources. Digitization may, in and of itself, stimulate improved access.

Digitization and the needs of local students and scholars:

Immediate demand can inject a measure of practical reality into decisions to create electronic resources. A historian may choose to teach from digitized images of manuscripts that would otherwise be unavailable to a large class. Because ready access to shared electronic files can transform the classroom, proposals to digitize in support of immediate teaching needs may garner faculty support.

Various approaches of Digitization and its facilitation to a researcher:

Different digitizing techniques result in electronic files with different characteristics. These in turn can correspond well or poorly with scholarly needs. If the goal is to provide an image-based finding aid that helps users identify original materials of interest, for example, mounting slow-loading high-resolution images would be counterproductive. If, on the other hand, the intention is to reduce or eliminate handling of original materials, an image that fails to convey all critical information embodied in the original will fail to serve its intended purpose.

The simplest approach to digitizing involves use of a scanner or digital camera to create electronic pictures (bitmap images) of original materials:

Decisions concerning the number of dots recorded by the scanner (resolution), how many shades of gray or colors will be recorded (bit depth), and other factors related to scanning equipment and settings will determine how well the digital product replicates the original. High-quality bitmap images can usually capture all the significant detail in texts or graphics. Scanning rare and unique texts or visual resources can make them accessible to users who would otherwise never see them. In such a case, merely reproducing the original in electronic form represents an extraordinary enhancement.

For textual materials, post-scan processing can support expanded capabilities:

Scanned text can be processed with Optical Character Recognition software to produce searchable indexes. OCR software is now only occasionally employed in digitizing projects because it cannot yet interpret accurately all fonts and alphabets, and because it adds significantly to per-page costs. Text can also be rekeyed to create ASCII files-very straightforward digital text files that permit searching by keywords or phrases. In some cases this enhancement is the primary justification for digitization. Directories, dictionaries, and indexes are all significantly easier to use when specific words can be searched within a well-designed digital file.

ASCII texts accommodate key-word searching (e.g., searching for all instances of the word "temperance") and some kinds of analysis, but they do not readily replicate the structure and format of an original document. Without special coding, researchers cannot directly consult the seventh paragraph of the third chapter of a particular text. Nor can they search for all occurrences of "welcome" used as a verb rather than a noun. These capabilities become possible in marked-up texts, which are coded to highlight elements of structure, format, and syntax. The Standard Generalized Markup Language (SGML) is the emerging model. One SGML

application, the Encoded Archival Description (EAD), is being used to create electronic versions of archival finding aids.

These and other approaches to digitizing carry very different costs, benefits, and resource requirements. While electronic versions can be more versatile than original materials, in some cases they hinder research. A scholar studying bookbinding or papermaking, for example, is poorly served by a reproduction of any kind. So too is the scholar whose immediate access to a large and important collection of literary works is sacrificed in order to serve a worldwide constituency-perhaps because bound volumes have been disbound for scanning.

Digitization increase the utility of the source materials:

Digitization can enhance original materials in many ways. Image quality can be improved by eliminating extraneous stains and marks. Thumbnail images of visual resources (photographs, drawings, and paintings) can be browsed to discover patterns, trends, and relationships among individual items, and specific images can then be scrutinized at higher resolution. Likewise, patrons can review scanned images to identify needed materials before requesting that they be retrieved from storage.

Electronic transcriptions of texts, in ASCII format or marked-up files can be linked to bitmap images of original documents. Readers can then decide for themselves whether "authoritative" transcriptions are in fact accurate. Comparisons of different versions of a text are likewise simplified. Related texts and images can be assembled together within a single, unified corpus. Examples such as the Geet Govind project of UNDP, administered by Indira Gandhi National Centre for Arts, New Delhi is an interactive, multimedia document on Archaic and Classical India, suggest the potential of electronic texts.

Almost all electronic products will provide basic links that allow users to navigate them (to locate a particular map within a printed text, for example). The

degree to which a digitization project exploits electronic links will depend upon its intended use. For digital resources created as pedagogical tools, predetermined connections are part of the package. Products intended for research tend to be less aggressive in ordaining relationships among sources, since their creators assume that researchers will build their own structures of meaning.

Critical features of the source material must be captured in the digital product:

The cost and nature of digitizing hardware and software continue to evolve, and preferred solutions are likely to shift as well. It may sometimes make sense to defer certain digitizing projects so that technology can catch up to needs. The success of a project to digitize oversized maps at Columbia University, for example, depended partly on the ability of users to see detail and read place names. As a result, the maps were scanned at relatively high resolution, thereby creating challenges for digital image delivery and presentation. File sizes were very large and initially outran the capacity of the library's computers and network. Greater bandwidth and more powerful machines have enhanced functionality.

Digitization process retained source material:

Automatic sheet feeders are fast and efficient, but they may destroy brittle paper. Digital cameras can minimize the manipulation of source materials, but subjecting certain media-watercolors, for example prolonged lighting is problematic.

Issues of hardware which is used for conversion:

Color slides, for instance, cannot be fully represented by scanners that create only black-and-white images. Even a color scanner with limited capacity to reproduce tonalities will be inadequate when high-quality images are important. Digitizing equipment can be expensive, and the costs may be difficult to justify when use is sporadic. Some projects may thus be done most economically if they are contracted out. Agreements with external vendors, in addition to specifying technical conditions,

performance expectations, and handling guidelines, must fully define ownership and distribution rights for all digital products.

Digitization of information resources which continue to grow:

Ongoing commitments and extended arrangements for copyrights may be required when collections are still expanding, as is the case with current journals and annual reports, or the papers of a living individual. Consultations with scholars and other experts can be particularly useful, since the long-term value of current materials is often difficult to discern.

Users navigation within and among digital collections:

Printed sources orient readers by means of tables of contents, chapters and sections, pagination, indexing, and formatting cues. Manuscript materials often rely on finding aids linked to the organization of file folders. Photographs may be mounted in albums. At a minimum, electronic products need to provide the same kind of functionality. The process may require several steps. For a multi-volume work that has been scanned page by page, for instance, each page is a separate computer file that must be individually labeled and stored. The files for critical pages of the work-for example, the title page, table of contents, and the first page of every new chapter-must then be linked to electronic navigational tools so that they can be easily located.

Existence of Digital files in the Information Centers:

Bibliographic records, finding aids and indexes can all be adapted to include references to electronic resources. Nonetheless, our ability to determine what has been digitized remains well behind what we know about materials that have been microfilmed or photocopied.

One of the principal challenges is to determine what information is essential in describing an electronic product. The "Dublin Core" and other special initiatives for structuring and standardizing descriptive data propose to combine information about

the technical characteristics of digital files, their location, and a summary of their contents. The resulting records are known as "metadata." Their function is to provide users with a standardized means for intellectual access to digitized materials. Despite these and other initiatives, projects to catalog digital files are only in the developmental stage. No system has yet been widely adopted for tracking the digitizing activities of libraries, although new approaches continue to emerge.

Better delivery of Digital products to users:

Alternative modes of digital storage and delivery must be considered from the outset of a project. CD-ROMs, for instance, are distributed and used differently from information made accessible over the Internet. The differences are reflected in hardware requirements, software, and ease of use. CD-ROMs are sometimes bundled with software for searching and analysis that is superior to that generally provided for Internet files. On the other hand, access to CD-ROMs is limited to individual workstations or small networks, while Internet files can be made available to a very broad audience. And Internet resources, by nature, can be updated or augmented without requiring users to replace objects that have become obsolete.

Internet products, however, generate questions of their own. How immediate must access be?

Files can be mounted on a server so that they are instantaneously available on-line. They can be stored on disks in a jukebox and loaded on demand ("near-line" access), or kept off-site ("off-line") and retrieved and delivered on demand. Near-line and off-line access can save on server space and requirements, though there are countervailing staff costs associated with retrieving and mounting the files. Expected demand, file sizes, fee structures, and available staffing and equipment must all be considered.

Authorization of use of the digital resources, and required circumstances:

Copyright holders may limit distribution rights, institutions may be unable or unwilling to provide the infrastructure needed to support universal access, and cost-recovery enterprises cannot by definition make their products available without restriction. Digitizing projects must thus consider access policies and control, pricing mechanisms, and billing procedures. Access issues impinge upon selection decisions in a number of ways. A university may mount high-resolution images of unique holdings for scholarly use (a medieval manuscript, an important collection of drawings), but would not allow unauthorized publication of those images. Moreover, electronic resources cost money that must be secured through subsidies or fees. When neither internal budgets nor external subventions provide adequate financial support, digitization will require a paying audience.

Access, when it is not universal, must be managed:

Current alternatives include passwords, direct user fees, and limitations according to organizational affiliation. Different capabilities for viewing, downloading, and printing may be offered at different prices or to different sets of users. There are many options, each reflecting a different pathway toward a self-sustaining endeavor.

Ensure the integrity of the digitized data:

The malleability of electronic products makes them particularly useful for many kinds of scholarship. Digitized files must be embedded with detailed information concerning the methods used to create them. The same information should be included in external bibliographic or descriptive records. Users who are consulting or copying the sources must also be able to confirm that the files they see or receive match the originals. Means to authenticate and protect digital products, long available in financial and industrial applications, are only beginning to take hold in the scholarly world.

Adequacy of the existing technology infrastructure to fulfill local demands:

Robust computer systems and an appropriate number of work stations are perhaps more easily provided than such ancillary features as network printing capabilities in the library and in offices, classrooms, and residences.

The Goal of the project is to long-term preservation of deteriorated materials:

Preserving documentary resources in electronic format presumes that, to the greatest extent possible, all the information contained in the original material has been captured completely and accurately. This requires careful attention to significant detail, whether the smallest text characters on a page or all the shades and tones of blue and green in a seascape. Targets for resolution, grayscale, and rendition of color either exists or is being developed to ensure the needed detail and fidelity.

Digital preservation also requires a supporting organization and infrastructure dedicated to storing the electronic files and to migrating them to new formats and/or media as technology changes. Unless these capacities are all in place, digital files cannot be regarded as permanent. Creating an enduring digital preservation master file is a multidimensional task with long-term implications. Hybrid projects, in which digital files are complemented by copies on microfilm, alkaline paper, or some other stable medium, provide the insurance that exclusively electronic projects do not.

Availability of already digitized source material:

As we have seen, it can be difficult to determine whether a specific item has been digitized and by what means. If an electronic copy does exist, is it accurate, satisfactorily functional, and accessible? Does it take advantage of the capabilities of current technologies? If the existing product does not serve the intended purposes of the proposed project, a new version may be warranted.

Benefited User Group of the proposed digital product:

It is important to consider whether the product will support better teaching or research and enable students to learn more, or in different ways-if, for example, texts or images are more fully revealed. Digitization may allow librarians to manage collections and provide services more effectively, or to provide traditional services such as copying or interlibrary loan at lower cost or at less risk to collections.

Commensuration of the intellectual value of the proposed product with the expense:

The limited resources available for digitization might have greater impact if they were directed at another project, or directed toward an entirely different approach to providing access-through exhaustive indexing perhaps, or microfilming, or some other type of reformatting that would prove in the end more useful to scholars.

Creation of acceptable digital product at lower cost:

When materials are scanned to support short-term course work, for example, careful (and expensive) post-scan processing to eliminate extraneous marks and speckles or to deskew misaligned images may be a waste of time. Likewise, an adequate substitute for full-text scanning of little-used journals might be provided by linking scanned tables of contents and indexes to bibliographic records and relying on traditional forms of document delivery.

How will the proposed project address the long-term costs associated with digital files?

The accumulated body of digital products may enable savings elsewhere in the institution-for example, by reducing staff costs for reshelving bound journals, or by lowering the costs of storage, circulation, and preservation-and these savings could offset some or all of the expense of digitizing. But such savings as may be realized are difficult to predict. It is essential to realize that the costs of digitization are just beginning at the time of initial capture. The programmatic capacity to distribute and maintain electronic resources, and to migrate them to new forms as original digital

platforms fail and formats and software are superseded, is fundamental to long-term efforts. In addition, there are staff costs associated with training and user support. Finally, rising user expectations may require that existing digital files be reprocessed in new ways. When OCR software is perfected, for example, unsearchable bitmap images of texts could be thought unsatisfactory. Projects that do not plan for change may become obsolete, and therefore irrelevant.

Research libraries are eagerly embracing the digital world:

They are acquiring access to great quantities of electronic materials produced outside their walls. At the same time it becomes essential to maintain digital versions of all worthwhile existing text material of their own holdings. A careful review, analysis, and planning can yield electronic resources that are functional and faithful to the original sources, and that support new kinds of scholarship. A detailed plan of work, regular assessment of progress, closely documented adjustments and corrections, and the retention of other project-related data can strengthen the knowledge base for future efforts.

4.94 ESCRIBING AND RETRIEVING PHOTOS USING RDF AND HTTP:

4.941 Introduction:

Describing & retrieving (digitized) photos with Resource Description Format (RDF) metadata describes the RDF schemas, a data-entry program for quickly entering metadata for large numbers of photos, a way to serve the photos and the metadata over HTTP. The data-entry program has been implemented in Java, a specific Jigsaw frame has been done to retrieve the RDF from the image through HTTP. The RDF schema uses the Dublin Core schema as well as additional schemas for technical data.

Diagram of the parts of the photo-RDF system. Top left: the pictures are digitized and stored as JPEG images. Bottom left: metadata is written into the pictures with the

data-entry program (and can also be edited if corrections are necessary). Right: requests from the Web are served by Jigsaw, by sending either the picture or the metadata, depending on the form of the request.

The system comprises the following, largely independent, pieces:

- Scanning the photos and storing them in JPEG format. Negatives are good for scanning the best quality, but any process that yields JPEG could be used, including digital cameras.
- A data-entry program that allows easy entry/editing of the metadata for each photos and stores the data in RDF form inside the JPEG file.
- A module for the Jigsaw server that can serve either the JPEG image data or the RDF description that is stored in it, using HTTP content negotiation to determine which of the two a client wants.

Some digital cameras are already producing information about the picture, which may be read and reformatted in RDF by scripts. The RDF data is expressed in three separate schemas, one of which is the Dublin Core schema. The other two deal with technical data of the photo and with subject categories. The reason for using three schemas is solely to allow each of them to be used in other projects, to the users of the data-entry program the actual RDF is completely hidden.

4.942 The data-entry program "rdfpic":

Screen dump of **rdfpic**, the metadata editor, showing the screen to enter technical data. The data-entry program is very simple. It has been designed to enable quick entry of metadata for lots of photos, under the assumption that the photos will usually be from one or a few series. Most fields therefore show by default the value that was entered for the previous photo, and give quick access to the values entered for the last few photos. Typically, only very few fields will have to be changed from one photo to the next and the amount of typing will be minimized.

The program is written in Java, but the user interface is in fact generated at run-time directly from a machine-readable version of the schemas (currently not the RDF syntax, but a transformation of it, with equivalent information). This means that the program does not need to be changed when we change the RDF schemas.

The RDF data is stored in the JPEG file in comment blocks (blocks of type "COM", as defined by ISO DIS 10918-1). According to the JPEG standard, a comment block can contain arbitrary text. There is no way to assign a type to the text. It is a matter to rely on the fact that RDF can easily be distinguished from plain text by heuristics. JPEG limits each comment block to 64K, but there can be as many blocks as necessary, so arbitrary amounts of text can be added. In practice, the descriptions generated by the rdfpic program are typically only a few hundred bytes long.

4.943 The Jigsaw extension:

To serve either the RDF version or the complete image using existing browsers and tools, the best way is to use Content Negotiation. It doesn't exclude the use of other techniques, such as HTTP extensions, to be able to retrieve and store metadata in a better way.

Using Content Negotiation it will provide following benefits:

- It will work right away with all text-based browsers (lynx, emacs with emacsspeak, etc.) and
- The output can be rendered directly by selecting, e.g., the title or the description from the RDF.
- An RDF crawler will be able to get all the descriptions of a collection of photos to create a knowledge database, just by asking for the right MIME type.

In Jigsaw, a frame has been created, to simulate two different resources under the same URI, the one of the image itself. Those two resources have their own set of HTTP

values, such as ETags, Content-Length and others and the result is sent out using the classic Content Negotiation of HTTP. The RDF can be also be fetched directly without doing Content Negotiation, by just adding the wanted MIME type after a semicoln (;) e.g.: foo.jpg;application%2Frdf+xml ("%2F" is "/", escaped for occurrence in a URL.)

4.944The RDF schemas:

The metadata is separated into three different schemas:

4.9341 Dublin Core schema: The Dublin Core schema is a general schema for identifying original works, typically books and articles, but also films, paintings or photos. It contains such properties as creator, editor, title, date of publishing and publisher. The Dublin Core Metadata Initiative is developing it and the version of our interest is the RDF-format of version.

4.9442 Technical schema: This schema captures technical data about the photo and the camera, such as the type of camera, the type of film, and the date the film was developed and the scanner and software used for digitizing.

4.9443Content schema: This schema is used to categorize the subject of the photo by means of a controlled vocabulary. This schema allows photos to be retrieved based on such characteristics as portrait, group portrait, landscape, architecture, sport, animals, etc.

All the properties are optional. The more properties are given values, the better the photo will be described and the easier it will be to find it, but leaving properties undefined doesn't make the metadata invalid. There are no dependencies between the properties, each property can be given a value independent of whether any other property has a value. The values are also independent, except for restrictions of common sense.

4.9441 The Dublin Core Schema:

Here is an interpretation of the Dublin Core properties, applied to photo material. In parentheses the label that is shown in the user interface of rdfpic, if it is different from the property name.

Title: A short description of the photo.

Subject: A set of keywords to describe the photo. See the content schema below for the list of keywords.

Description: A longer description of the photo.

Creator ("author/creator"): The photographer, as a URL that can be further described with other schemas.

Publisher: The person or institution making the photo available, often the same as the creator.

Contributor: A person who contributed in some way, e.g., the person who digitized the photo; may be a URL or a name.

Date: The date and time the photo was taken, conforming to ISO format [ISOdate]. The year is required, everything else can be omitted: yyyy[-mm[-dd[Thh:mm[:ss[.sTZD]]]]]. The default time zone is UTC. Example: 1999-10-01

Type: Always "image" (see the Dublin Core's List of Resource Types

Format: Always "image/jpeg"

Identifier ("number"): A number for the photo that is meaningful to the publisher. This is not the URL of the photo and it does not have to be globally unique.

Source: not used.

Language: not used.

Relation: Identifies a series: the event or topic for a series of photographs. Can be a URL or a string.

Coverage ("location"): The location shown on the photo. (Note that we only use the "spatial coverage," not the "temporal coverage," since we assume that a photo is instantaneous and thus the date field is enough.).

Copyright statement, or the URL for one. Example:

http://www.example.org/People/Lafon/Copyright?1998

4.9442 The Technical Schema:

The technical schema is defined by this RDF schema:

Camera: The brand and type of the camera, or a URL for the camera. If the latter, the URL identifies one actual camera, not all cameras of that type.

Film: The brand and type of film. In contrast to the camera property, this is not an individual roll of film, but identifies all films of the same type. (We assume films of the same type are sufficiently similar; except for fabrication errors, they are interchangeable.) The value may be a string or a URL that is further described elsewhere. As a convention, digital cameras should be considered as "digital" film.

Lens: A definition of the lens used, maybe a URI describing it, a URI pointing to the camera for compact cameras, or just plain text description.

Date: Date on which the film was developed. The date must be in the same form as the date property. Example: 1998-08-04

4.9443 The content schema:

The content schema contains the keywords we use in the "subject" property of the Dublin Core schema. That property should contain as many of the following keywords as are applicable. The keywords have the following meaning:

Portrait: The photo contains a portrait of one person.

Group-portrait: The photo contains a portrait of a group of people.

Landscape: The photo contains a landscape or skyline.

Baby: The photo contains a baby.

Architecture: The photo contains interesting buildings.

Wedding: The photo contains scenes from a wedding.

Macro: The photo contains an extreme close-up and would, when viewed under normal circumstances, be larger than life-size.

Graphic: The photo contains a pattern, texture or design, that is interesting for its abstract, graphic quality.

Panorama: The photo contains a wide-angle view of a landscape or skyline.

Animal: The photo contains an animal.

4.95 DOWNLOADING THE CODE:

The Jigsaw extension and the JPEG related classes are a available in the Jigsaw </Jigsaw/> 2.0.4 distribution, the metadata editor rdfpic <http://jigsaw.w3.org/rdfpic/> is available from the Jigsaw demo site <http://jigsaw.w3.org/>.

4.96 CONCLUDING REMARKS:

A digitization project can cover a wide range of complex activities and it is often easy to lose track of the underlying project aims and objectives. Digitization is a tool and not a purpose and should always be used to facilitate the end result of the project rather than becoming the sole focus of it. It is hoped that this document will help to make the process of digitization less fearsome and more tangible and therefore something to be harnessed to help to create useful and exciting digitization projects.

CHAPTER 5 : TELECOMMUNIACTION TECHNOLOGY

- TELECOMMUNICATION IN DIGITAL LIBRARIES

- WEB SERVER TECHNOLOGY

5.1 MODES OF TELECOMMUNICATION IN DIGITAL LIBRARIES:

Digital libraries or digital collections are becoming ubiquitous in the information arena. Certainly the most fascinating and challenging with regard to access are those, which are predominantly or solely image based. A prime example of such collections is the American Heritage at The Library of Congress. Improvements in computer and telecommunications technologies have enabled information professionals to include them in their offerings either internally within the organization or as links to external sites. The need to access such collections is not only vital to basic research, but also invaluable to human communication in the digital age. The adage "a picture is better than a thousand words" has never been more appropriate as when it is used to refer to digital resources on the Web. The Internet, particularly the Web, has made it possible to access such collections and has in fact accentuated the creation of remotely accessible image-based digital collections. But unlike text-based information access, image intensive digital libraries are fraught with downloading and uploading bottlenecks. Careful design of distribution and receiving information systems is needed. Various alternatives have been used to alleviate the bandwidth bottleneck, including: cable modems, frame relay, ISDN, digital subscriber line (DSL), satellites, and high-speed analog modems. Comparative analysis of various alternatives of Digital library access is presented along with the information of their development expenditure and incoming/outgoing performances;

Table I Alternatives for digital library access

Technology	Deployment	Monthly cost ($)	Performance incoming	Performance outgoing
A-modems	Universal	20	56kbps	33.6kbps
ISDN	Widespread	50-130	128kbps	128kbps
C-modems	Limited	30-65	1-5mpbs	33.6kbps-2.5mbps
DSL	Limited	49-1,200	144kbps-9mbps	64kbps-9mbps
Satellite	Widespread	40-130	400kbps	33.6kbps
F-relay and T... series	Widespread	300-3,000	56kbps-45mbps	56kbps-45mbps

Alternatives for digital library access

It summarizes the latest developments in these technologies and how they can be used by various types of professionals and end-users in accessing digital libraries. While in a given information systems environment, the mode of implementation of the transmission technologies in the physical networks significantly affects the ease with which digital libraries are accessed, the basic access terminal has also to be reckoned with. Thus three major elements interplay to determine the final rate at which multimedia in the digital library are effectively transmitted to the end-user. First, an effective national telecommunications network, that includes signaling and switching techniques must be in place and operating efficiently. Second, an institutional distribution network which delivers information to the end-user must be available. Third, the access terminal used, which for many end-users is currently an intelligent personal computer (PC) must have the capacity to handle a variety of images and sound. Assuming that the intelligent PC is readily available to the end-user, albeit in a variety of flavors and degrees of regional penetration per capita, the telecommunications networks become the more significant elements.

5.11 The Communication Technologies:

5.111 High-End Analog Modems:

Starting with the lowest on the totem pole, Plain Old Telephone Service (POTS) as the delivery network for digital images, we have an analog modem based access system with maximum delivery speed of 56kbps. This is slow for multimedia and is certainly inadequate for teleconferencing, video conferencing, and animation, which may tap into digital libraries for source images and sound. However, the majority of Internet (Web browsers, especially in homes, still use analog modems to access digital libraries. There are several advantages of using analog modems. First, with a typical lowest recurring cost, they are affordable for small library and users. Second, they are part of a universally available POTS network and thus accessible to most users. Third, they are easy to install and in many instances come already installed in new deliveries of computers. Finally, they are easy to maintain and operate. The greatest disadvantage of analog modems is definitely speed for downloading and uploading of digital images,

which eliminates them as viable alternatives for optimum data delivery in medium and large organizations.

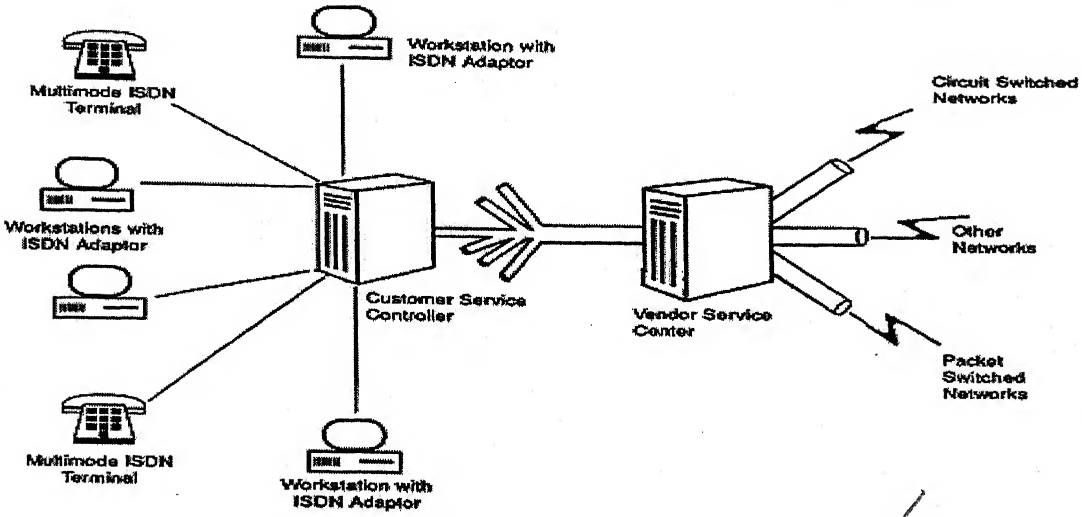
5.112 T... Series And Frame Relay:

Since the technologies in this section are very complex and expensive they are dealt with here to dispose of them as viable alternatives for most end-users or small to medium organizations. Some of the critics question on their cost benefit advantages even for large organizations, especially in their small branches for telecommuting employees that may have access to cable modems, satellite, or DSL. T... series (i.e. T-1, T-2c T-2, T-3 T-4) are telecommunications standards that define long distance digital lines used mainly for data communications. AT&T, and several North American communications carriers own T... series lines for leasing, many of which are fiber optic-based and digital. Their transmission speeds range from 1.544-274.176Mb per second (mps). Many large organizations lease dedicated or shared fiber optic-based digital T... series trunk lines for their own proprietary information networks such as metropolitan area networks (MANs) or intranets. They are very fast and ideal for accessing distant digital libraries - if only they were affordable. The main disadvantage for T... series are the charges, which may range from $300-$3,0000 or more per month. Frame relay is another method of data transmission with transmission rates ranging from 56kbps-1.5Mbps, but within the same price tag as the lower end of the T... series.

5.113 ISDN In Digital Transmission:

To augment the POTS, ISDN was one of the earliest broadband signaling systems developed in the 1980s, by the telecommunications carriers. It does indeed transmit images at a faster rate than the analog modem-based networks. The ISDN signaling algorithm works on the regular telephone network and requires ISDN switches at the telephone company's central office and an ISDN capable terminal at the user end. As it establishes a virtual digital network, it achieves high efficiency, for there is no signal conversion, comparable to the analog POTS, at either end - carrier or subscriber. Its two main flavors, the basic rate interface (BRI) and the primary rate interface (PRI) carry

signals at maximum rates of 128kbps and 1.544Mbps respectively. Such rates are a great improvement on the POTS analog modem-based digital image transmission.



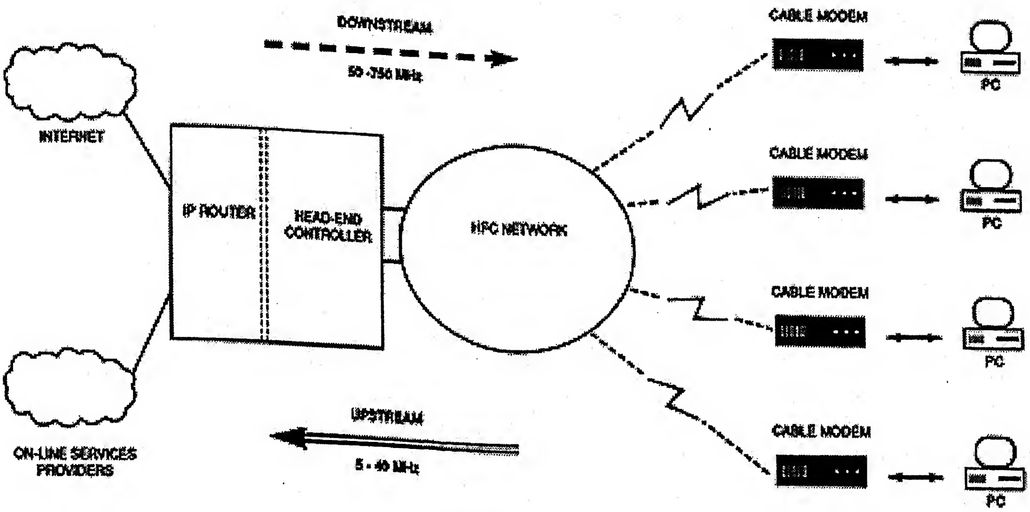
Digital library access via ISDN

Although it has been on the market for almost 20 years as an alternative technology for transmitting digital images, ISDN has been slow to develop. Among the reasons often given are cost of equipment and installation, uneven deployment, and lack of "trigger" user application at its critical developmental period during the 1980s. Even in countries like the USA, with highly sophisticated telecommunications infrastructures, ISDN has not taken off as anticipated by the telecommunications carriers. Internet user demand for high bandwidth to accommodate digital images, as well as the threat of other technologies such as cable modems, accounted for the heightened telecommunications company's interest in ISDN in the 1990s. But by the time ISDN became affordable, alternative technologies like DSL and cable modems were on the market and offered a better multimedia operational environment with higher bandwidth and faster throughput.

5.114 Cable Modems In Digital Transmission:

Cable modems are one of the most promising technologies for accessing digital libraries for the twenty-first century. The monthly cost for the technology is lesser then others. Most of the deployed systems are based on the existing CATV networks, which are plagued by two intrinsic design problems when modified to accommodate bi-

directional data communications. First, they were never designed with a symmetrical two-way point-to-point communication mode in mind. Second, they have been deployed on a regional or local basis with virtually no common standards and little quality control. Ordinarily, television programs are broadcast down the cable simultaneously in what is called a one-to-many mode, normally called the downstream. Because of the differential in the volume of signals upstream and downstream, most CATV companies use a split frequency spectrum for communication in either direction.



Digital library access by cable modem

The downstream is allocated 50-750MHz of the spectrum, which is the regular analog television broadcast band. This band is traditionally a one-way, or technically a simplex, method, which over the years of technical experience has been perfected in terms of error checking, trouble-shooting, and signal amplification requirements. Depending on need, digital data or video may be sent downstream to users (in homes or offices) using one or more of the unused channels within the broadcast band.

The upstream band used for data communications between the user terminals (PCs) via cable modems to the head-end is allocated the 5-40MHz part of the spectrum. Thus, the upstream portion of the network uses a many-to-one data communications mode, whereby data is sent to the head-end controller by the individual nodes on the network. It has a lot of noise. usually known as ingress. Such noise is external to the

network and comes from home equipment such as dryers, mixers, thermostats, as well as ham radios and is caused by the low frequency on the upstream channels. Cable network providers are developing filters to minimize such noise.

At the users level, cable modems tune to relevant channels, demodulate the signals transmitted from the head-end and sends them to the users' personal computers. It reverses the process on the return circuit. This simplified picture of the cable modem needs some elaboration, for it has more functionality than a traditional telephone modem. The cable modem also acts as a transceiver, i.e. it receives and transmits data to and from the head-end. In addition, depending on sophistication, it may contain routers, and diagnostics management software. For Internet access, each cable modem has an Ethernet port, which facilitates connection to the computer on one side and the cable connection on the other. The user must install an Ethernet adapter inside the PC, and connect it to the cable's Ethernet port by an RJ-45 connector. Appropriate software is used to configure the PC to operate the TCP/IP protocol and make a direct connection to the Internet. Depending on the cable network connected to, maximum downstream speeds may be 500kbps-30Mbps, while upstream may have 96kbps-10mbps.

Deployment of cable modems excites many people who are struggling with slow uploads and downloads, although cable modems do have inherent problems. First, many of the cable networks are not two-way capable and some analysts suggest that only 5 per cent of cable networks can deliver broadband without major upgrades. Second, one's neighborhood may not be among those served by broadband bi-directional cable networks. Third, standards are lacking and if a customer moves from one region served by one cable provider to one by a different company, there is no guarantee of compatibility. However, IEEE P802.14 Cable TV MAC and PHY Protocol Working Group, and ATM Forum's Residential Broadband Working Group are working collaboratively to alleviate the lack of standards problem. Finally, cable modems are still quite expensive compared to telephone modems.

5.115 DSL In Digital Transmission:

Digital subscriber line (DSL) is a relatively newcomer to the telecommunications market, but perhaps the most promising to transmit digital images to the end-user in corporations as well as homes. Asymmetric digital subscriber line (ADSL) is one of the flavors of a group of several digital signaling techniques that have been developed in the last decade or so to utilize the existing telephone network to carry high bandwidth. These systems go by digital subscriber line as the generic name and are sometimes collectively treated together as xDSL, where "x" is a variable replaceable by any specific character for the particular type of DSL. Other examples from the family of DSL signaling systems include: Symmetric DSL (SDSL) and Very high rate DSL (VDSL), which differ in both the mode of transmission and bandwidth. The renewed interest in DSL is attributable to the Internet demand particularly for the Web-based large files, especially digital images.

Since it is based on a signaling algorithm that uses the regular telephone twisted copper wire network, the very foundation for POTS, DSL has a nation-wide appeal for transmitting multimedia. Some analysts have asserted that it holds the greatest potential for mass deployment as it introduces the broadband characteristics needed for high volume image-based digital data transmission on a network, which is associated with a conventionally narrow band signal transmission. ADSL's asymmetric mode of transmission is well suited for Web multimedia access. Most users request digital images from remote servers using few textual commands, thus requiring minimal use of bandwidth upstream. At the server end, the downstream, massive multimedia is often requested requiring heavy use of the available bandwidth. ADSL is designed to serve such environments. The excitement with ADSL is justified, because the copper-based telephone network is the most ubiquitous in the industrialized world. Assuming that a reasonably stable POTS infrastructure exists, ADSL is poised to be one of the gems of the telecommunications industry for the twenty-first century. The often-quoted maximum transmission speed of 9Mbps is faster than analog modem speeds as well as ISDN. The use of satellites discussed in the following section improves access to digital collections especially in regions not well served by telephone or cable networks.

5.116 Satellites In Digital Transmission:

Communications satellites are an alternative medium for transmitting multimedia. They are ideal for sparsely populated areas or areas which have not been adequately covered by the regular telephone or cabling networks due to adverse terrains. They are also the main technology of choice for linking less-developed countries to advanced countries databanks for digital image access. Comparable to microwaves, their mode of transmission is based on high-frequency radio waves with very high bandwidth. Their mechanism includes a space satellite and two or more ground stations. The earth stations used for multimedia communications are similar to dish antennas commonly used by individuals or organizations to receive television signals. Two typical terminals have characterized satellite digital data access at the end-user end. Very small aperture terminals (VSAT) are mainly for text, while T-carrier small aperture terminals (TSAT) can carry multimedia as they achieve a 1.544Mbps data rate. Most communications satellites are placed in a geostationary orbit - an orbit timed to the earth's rotation approximately 23,000 miles above the earth's surface. Within such an orbit, the satellite stays in a fixed position with regard to the earth's antennas. This obviates the need for constant re-orientation of the earth's stations in order to remain in touch with the communications satellite.

Hughes Network Systems DirectPC is one of the pioneer systems developed with telecommuters, the general home end-user, and small to medium corporations in mind. One advantage of satellite digital image communications is the high bandwidth, which is ideal for digital image transmission. Satellite digital transmission mode is essentially broadcast in nature. This implies that messages beamed to the earth may be picked up by any station tuned to a given radio frequency channel and pointed to the communications space satellite. Although this allows the satellite to send signals to many earth stations simultaneously, within its footprint, privacy of data is hard to maintain. For corporate or otherwise confidential multimedia data, scrambling or encryption is normally used. At the receiving station, such data must be deciphered using special conversion algorithms. Yet another serious problem with satellites is propagation delay caused by transmitting

signals through space. All satellite signals using a relay station in ordinary geosynchronous orbit are subject to a quarter of a second delay in both directions. While the delay may be vital to some real time interactive data applications, file transfers can be done with relative convenience.

Given the exponential growth of the Internet and other information networks especially the World Wide Web, access to digital collections containing graphics, sound and moving pictures has become imperative. Information systems designers and end-users have to grapple with downloading and uploading bottlenecks, which are gradually being solved by using emerging telecommunications technologies.

While comparatively analyzing the costs of the relevant technologies using "lowest mean" costs as the baseline. While frame relay and the T... series (T1 ... T4) are used and afforded by giant corporations, small businesses, telecommuters and or other home users, and users in isolated areas may select one of the newly emerging technologies: DSL, cable modems, or satellite, depending on location. Competition in the last few years has definitely lowered the costs. Uneven deployment is also true in other industrialized countries where some of these technologies are marketed.

5.2 WEB SERVERS:

5.21 APACHE WEB SERVER:

5.211 Introduction:

The Apache Web server is the crown jewel of the open source software movement. It costs nothing to obtain, performs better than the competition, and is thus more widely used than all other Web servers combined. The propagation of open source software is tightly analogous to biological natural selection-the Linuxes and sendmails of the world eventually end up on the cover of Time magazine and are swallowed by the hype machine, while legions of DOS utilities slide slowly but inexorably to the /dev/null of history. Apache would not be popular if it didn't work well.

Apache has another virtue not quite so common in the open source world: It is simple enough that any reasonably competent computer user can master it. This is no slur on Linux, by the way; operating systems, particularly multiuser operating systems, are hugely complex. The only way to make them accessible to the average user is to dumb them down.

The collection of tasks delegated to Apache is thankfully not quite so vast. If somebody approaches Apache with little more than self-confidence and a sense of adventure, will be relieved to know that the configuration and care of the server itself really isn't a particularly complex task. The trick, depending on on the level of experience, will probably be to grasp the fundamental concepts of the operating system, learn the commands to make the machine do what you want it to do, and absorb the jargon.

The Apache server is descended from the httpd server created by Rob McCool at the National Center for Supercomputing Applications (NCSA). In 1995, httpd was the most popular Web server in existence, but when McCoolleft NCSA in 1994, development of the program was stalled. A small group of Web administrators formed the core of what came to be known as the Apache Group. The members included: Brian Behlendorf, Roy T. Fielding, Rob Hartill, David Robinson, Cliff Skolnick, Randy Terbush, Robert S. Thau.

Together with contributions from Eric Hagberg, Frank Peters, and Nicolas Pioch, the Apache Group incorporated published bug fixes for httpd 1.3, added some new features, and released Apache 0.6.2 in April 1995. Since then, the Apache group, as they came to be known, has been fine tuning and enhancing the base software. Software ports are now available for virtually all the major operating systems, though the Unix platform remains the forerunner. The Apache Web server is the end result of an enormous coordinated effort by some extremely skilled programmers.

Apache exists to provide a robust and commercial-grade reference implementation of the HTTP protocol. It must remain a platform upon which individuals and institutions can build reliable systems, both for experimental purposes and for mission-critical purposes. It is believed that the tools of on line publishing should be in the hands of everyone, and software companies should make their money providing value-added services such as specialized modules and support, amongst other things. It is often seen as an economic advantage for one company to "own" a market-in the software industry, that means to control tightly a particular conduit such that all others must pay. This is typically done by "owning" the protocols through which companies conduct business, at the expense of all those other companies. To the extent that the protocols of the World Wide Web remain "unowned" by a single company, the Web will remain a level playing field for companies large and small. Thus, "ownership" of the protocol must be prevented, and the existence of a robust reference implementation of the protocol, available absolutely for free to all companies, is a tremendously good thing.

5.212 Open Source Software:

Apache is an open source product. Traditional shrink-wrapped software typically includes only the executable object code, not the human-readable source code from which it is compiled. Apache and the other open source products include with their distributions not only the executable object code, but also the source code files from which it was created.

From the end user's standpoint, this makes a lot of sense. For example, it is a common feature that the owner of a software may have a problem in his office. A large commercial software package running on a large commercial operating system may get into a state where it stopped responding to input and may be, in fact, unkillable. He may try a stack trace and a few other things, but without the source code, there really nothing to do. And the final solution is to dump out everything and shipped it off to the software vendor for analysis. Presumably, he'll get back to us in a week or two.

Apache and the other open source software products benefit from their constant exposure to the developer community. Because there are more developers working on each open source project than even the wealthiest corporation could afford to hire, flawed source code is located and fixed more quickly. The initial quality of open source code tends to be higher than that which was commercially developed. Because open source developers are motivated by the simple love of programming, you tend to get the best of the best working on open source software. Contrast this with traditional software shops, where much of the day is spent in meetings, on the phone, and trading stocks.

5.22 W3C'S JAVA SERVER (JIGSAW):

5.221 Introduction:

Jigsaw is W3C's leading edge Web server platform, providing a sample HTTP 1.1 implementation and a variety of other features on top of an advanced architecture implemented in Java. Jigsaw is a W3C Open Source Project, started in May 1996.

5.222 Different Jigsaw versions:

- **Jigsaw 2.2.2 (January 8th, 2003)**

  This new version fixes several bugs, and adds performance optimizations. It also provides HTTP compliance fixes. The only new feature is SSL support.

  - SSL Support for HTTP and WebDAV
  - HTTP/1.1 compliance
  - WebDAV support
  - Many bug fixes

This version fixes several bugs, including a security problem. It also provides new features: support for WebDAV in JigEdit, a PushCache package, and a validating filter

- HTTP/1.1 compliance
- WebDAV support
- PushCache package

    - (X)HTML validation on PUT
    - Apache mod_asis
    - Many bug fixes
- **Jigsaw 2.2.1(April 8th, 2002)**

- **Winie 1.0.8(March 9th, 2001)**

  Winie is a network utility to put files on the web using HTTP/1.1. The main feature of Winie is to solve the "lost update problem" ._Winie use the client side api of Jigsaw. Major changes are:

    - Content-Language support
    - Bugs fixed

  Common features of Winie are:

    - PUT, GET and DELETE files on the web
    - Version conflict detection
    - Retries when connection closed (like wget does)
    - Upload all files located in a directory (recursively or not)
    - Support for proxies
    - Support for metadata configuration (language, charset)

- **Jigsaw WebDAV Package (November 24th 2000)**

  WebDAV stands for "Web-based Distributed Authoring and Versioning". It is a set of extensions to the HTTP protocol, which allows users to collaboratively edit and manage files on remote web servers. This package is preconfigured as a WebDAV server.

- **Jigsaw 2.0.5 (June 5th 2000)**
    - Servlet API Support_ JSDK/2.2 support
    - JSP Support
    - Image metadata_extraction

- Many bug fixes.
- Jigsaw 2.0: It was developed by the World Wide Web Consortium (W3C), is designed to be a technology demonstration rather than a full-fledged release. It's purposely intended as a project to showcase new technologies, but in the case of Jigsaw 2.0, this Web server also ends up being more robust than the average Web server. Most importantly, though, Jigsaw serves as a useful blueprint to the future of the HTTP protocol and object-oriented Web servers.

5.223 Common Platforms:

The server will run on any platform supporting Java. At this time, it has been tested on Win95, WinNT and Solaris 2.x. Other people have reported successful use of Jigsaw on OS/2, MacOS, BeOS, Linux, AS-400 and AIX.

5.23 TOMCAT SERVER:

5.231 Introduction:

TOMCAT IS A Servlet Engine that operates the Java Server Pages (JSP) technique. The server side tech of Java becomes useful with Tomcat. The example infrastructure shown in the figure is based on the Tomcat Web Server, Java Server Pages (JSP), Java Servlets, Open Database connectivity (ODBC) and the MS-Access database. Tomcat web server includes a "build" mechanism, which separates the internal configuration details of the web server directories from the development area. The team can work on the various files for an application in a completely independent directory, and then "build" the application into the web server for testing and development. This approach also allows the developers to port the application from platform to platform, e.g. applications develop under Windows and deploy under Unix. The MS-Access database is also a good choice database for small and medium sized libraries although it lacks the sophistication of commercial products like Oracle.It is easily available with Microsoft Office and installed quickly; also, ODBC drivers supplied with Windows for

MS-Access. The best way to get familiar with the infrastructure is to work with the simple example provided with the Tomcat bundle.

5.232 Installing Tomcat Server:

Tomcat can be used as an add-on to an existing web server (currently Apache, IIS and Netscape servers are supported). A web application is a collection of resources such as jsps, servlets, html files, images, etc. which are mapped to a specific "URI" prefix. For example, all the resources related to OPAC database access are assembled into a "opac" folder and correspondingly all the requests that start with "/opac" can be mapped to this application.

The installation of Tomcat requires installing Tomcat web server and Java Development tools. Following versions of software are better to use for maintaining web access to OPAC database on LAN.

Tomcat	Tomcat 3.2.3
Java	Java 2 SDK 1.3.1
Windows	Windows-2000
Database	MS-Access

Tomcat and Java are installed, preferably on same drive at the specified location. The first and foremost task is to install JDK and Tomcat and get ensure that these are ready to run in a better mode. After successful installation the Tomcat server it will be essential to start Tomcat manually. To escape oneself for its manual starting, it is essential to configure as a Windows service and register for Auto option.

5.2321 Configuring the web application:

Before configuring it is essential to develop the web application in Java Server Page (JSP) to search OPAC database. It is essential to place all HTML pages and JSP applications in a new folder and images used in HTML pages are stored in the other folder. Before launching the application it require some configuration at server as well as on client site.

5.23211 Server Site:

While configuring the server site it is essential to create system Data Source Names (DSNs). These data sources are local to a computer but not user dedicated, any user with privileges can access a system DSN. One DSN entry is required for each database used for search. Each DSN entry includes database and its driver for providing the Open Database Connectivity (ODBC) to the source database.

5.23212 Client Site:

No configuration is required on client site.

5.24 INTRNET INFORMATION SERVER (IIS):

5.241 Introduction:

IIS 4.0 allows you to have multiple web sites on one machine. Though IIS 3.0 has this capability, IIS 4.0 expands the functionality of multiple web sites by adding additional characteristics to sub-directories, and allow for multiple applications. Special considerations need to be made when designing and administrating multiple web sites on a single machine, including when to use sub-directories, when to use virtual directories, how to handle security, and the handling of multiple applications.

5.242 Web Site Design:

Web sites should be singular entries that are self-supporting. Each web site should be able to be moved to a different machine for load balancing, or just transportation purposes. In order to do this, they should be self-supporting, have their own security, and their own application scope. If you are an Internet Service provider, you will want to be able to design, move, and have the user update their web site without interfering with other sites on the same machine.

The HTTP protocol uses URLs to request files from the web server. Since most of these files are contained on the file system, IIS needs to translate the URL to the full path

name of the file. The Internet Information Server does this translation on every request. However it is up to the administrator to configure the server so that the right URLs are mapped to the right directories.

To properly design the file system structure on a machine that hosts multiple web sites, it is necessary to understand the difference between a home directory, a virtual root, and a sub-directory. It is also essential to understand when to use virtual directories and when to use sub-directories.

5.2421 Home Directory:

A URL that just contains a domain name is requesting the home directory, sometimes called the root directory. For instance, the URL below is requesting the default file in the home directory.

The minimal amount of work that the administrator needs to do to assist the web server in mapping URLs to directories is to map the home directory. For example the home directory of this web site could be mapped to:

c:\inetpub\wwwroot

Now it is clear that what a home directory is, let us see how to create one home directory in IIS 4.0. The home directory is the starting location of the web site in IIS 4.0 and is created when you create a web site. IIS 3.0 didn't require a home directory, but IIS 4.0 does. Here is how to create a new web site and specify the home directory:

From MMC:

1. Select the server that you want to create the web site on.
2. Right Click and choose **Create New | Web Site**.
3. The New Web Site Wizard appears and you are asked to enter a web site description.
4. Enter a description and press **Next**.

5. The next page of the Wizard ask for the TCP/IP information, leave the default setting for now and press **Next**. You can always change these setting later.
6. The third page asks for your the home directory, enter in the directory and press **Next**.
7. The forth page queries you about the access permission. Select the proper entries and press **Finish**.

5.2422 Sub-Directories:

Sub-Directories are directories that inherit the URL mapping from the file system structure. For example, if this directory existed:

c:\inetpub\wwwroot\sales

Then this URL would also exists:

http://www.myserver.com/sales

Sub-directories do not need to be defined to the web server by the system administrator. Because of this, just creating the sub-directory with Explorer will create the directory. There is no need to make any modification in the IIS 4.0 configuration.

5.2423 Virtual Roots:

Virtual directories are sub-directories of a URL that are mapped to file system directories that might not inherently exist on the file system. For example if you wanted your site to contain the following URL:

http://www.myserver.com/marketing

And this directory didn't exist:

c:\inetpub\wwwroot\marketing

One can create a virtual directory that mapped the URL to:

c:\inetpub\marketing\website\external

Virtual directories make the web site appear as if it has a different directory structure than it actually has on the file system. Here is how to create a virtual directory in IIS 4.0:

From MMC:

1. Select the web site that you want to create the virtual directory in.
2. Right Click and choose **Create New | Virtual Directory**.
3. The New Virtual Directory Wizard appears and you are asked to enter an alias to the virtual directory.
4. Enter an alias and press **Next.**
5. The next page of the Wizard ask the physical directory location of the virtual directory, enter the physical directory information and press **Next.**
6. The third page queries you about the access permission. Select the proper entries and press **Finish**.

In IIS 3.0 the difference between virtual directories and sub-directories was significant. In IIS 3.0, sub-directories inherited the properties of the parent directories and virtual directories could have different properties. For instance if you made the home directory read-only and you created a sub-directory called scripts - that sub-directory would be read-only also. If you wanted the scripts directory to have read and execute permissions so that you could run ASP files, you would need to make it a virtual directory.

In IIS 4.0, sub-directories inherit the properties of the parent directory upon creation, but these properties can later be changed. In IIS 4.0 you can create a sub-directory called scripts and change its properties so that it has scripting permission without creating a virtual directory. Here is how to change the permissions of a sub-directory.

From MMC:

7. Select the sub directory whose permissions you want to change.
8. Right Click and choose **Properties** from the drop down menu.

9. The **Properties for** Dialog appears.
10. Choose the **Directories** tab.
11. Select the proper permissions and press **OK**.

Virtual directories should only be used when sub-directories can not be used. Here is where we get into personal opinion. Because sub-directories take no web server, and they have all the functionality of virtual directories in IIS 4.0, they should be used whenever possible. Plus sub-directories organize all files into a central location for the web site.

Virtual directories should be used when all the files in the virtual directory does not fit on the physical disk. For instance, if you have a web site that is bigger then 2 Gigs, you might not be able to fit all of it on one disk. In this case, you will need to separate the web site into multiple virtual directories on the directory on each disk. For performance you can also divide your web site up onto multiple disks. In theory, random access across multiple disk drive should be faster then the same number of accesses on the same disk.

If you have multiple web sites and you are sharing information, virtual directories can be used to accomplish this task. For instance if you are sharing graphics, both web sites could have a virtual directory called graphics that is mapped to the same physical disk location. This would be impossible to do with sub directories. Updates to the files in the graphics directory would effect both sites. There is also a performance consideration here, two sites sharing the same files would allow NT to do more memory caching of those files than if they where in separate directories.

One of the main differences between IIS 3.0 and IIS 4.0 is Application Scope. In IIS 3.0, the scope of the application covered the whole machine. In IIS 3.0, if you had two web sites running on the machine, they both shared the same application. In IIS 4.0, you can have more then one application in each web site, and many applications scopes

on the whole machine. In order for each web site to be a singular entity you need to understand how to assign each web site it's own application scope.

For instance, in IIS 3.0 if you have the web site http://www.myserver.com and http://www.myofferserver.com and the user linked from one of the servers to the other, he would be within the same application scope. Which means that if you had two global.asa files, one for each web site, only the first global.asa would be called and the second would not. The one called would be the global.asa that corresponded to the first web site that you entered on that machine.

IIS 4.0 gives you the ability to have an application scope start anywhere that you have a directory. The scope then extends to all files in that directory and all files in the subdirectory below. The subdirectory rule however only pertains if there isn't another application scope defined in any of the subdirectories themselves.

One of the problems with the word security is that it means different things to different people. In this issue we will be referring to security as the ability to restrict access to pages on the web server. In the IIS context, security can also refer to SSL encryption, which we will not be addressing.

Most of us run anonymous security configurations and do not think much about web security. However, if you are going to secure your web site, you will want to design the file structure to make administrating security easy. You also need to take into consideration multiple web sites on the same machine, each might have different security requirements that need your attention.

CHAPTER 6 : DATA ANALYSIS

6.1 Introduction:

As it is mentioned earlier in the scope that selected open universities will be the part of the study, only four open universities have been surveyed, these are:

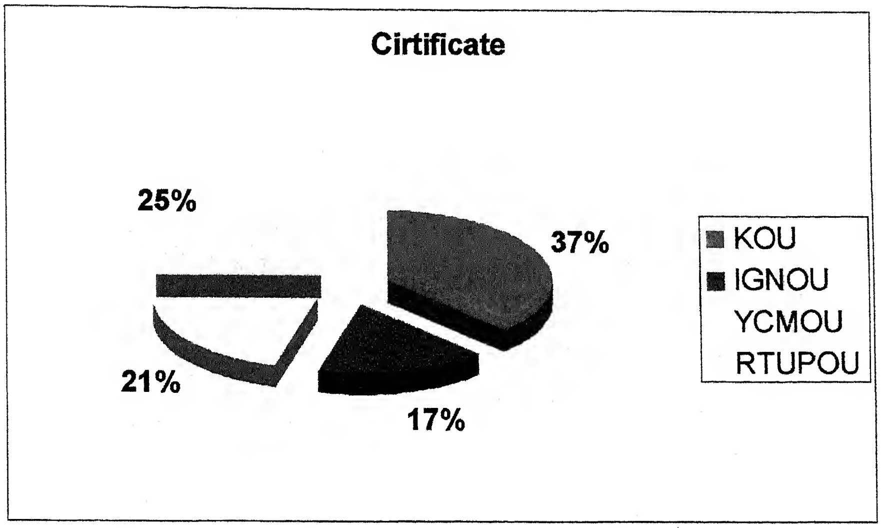
Kota Open University(KOU), Kota (Raj.)
Indira Gandhi National Open University(IGNOU), New Delhi
Yashvantrao Chavan Maharashtra Open University (YCMOU), Nashik (Mh.)
Rajarshri Tandon Open University, Allahabad (U.P.)

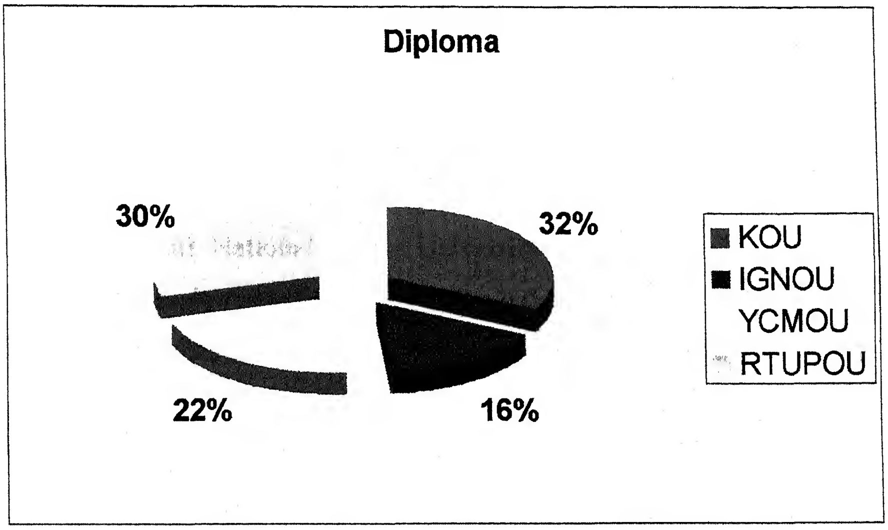
The population surveyed in all of these open universities is 50 open learners, which have been selected randomly from study centres and the information desk of the university head quarter. Tools used during the survey are well designed questionnaire for the open learners and the interview schedule.

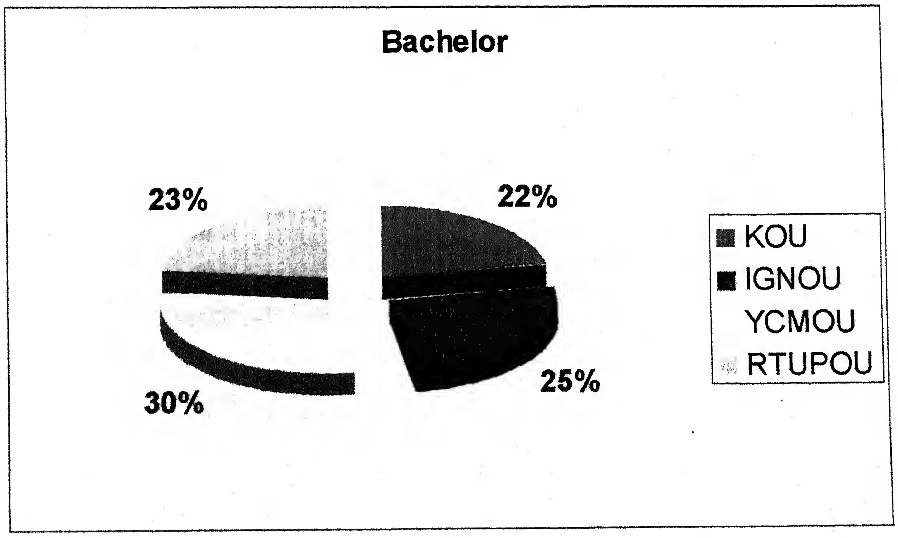
Data is presented in the form of various categories, identified at the lime of literature survey and designed in the form of questions as the part of the questionnaire. The questionnaire is distributed among open learners. The collected data is processed and presented in terms of pie chart of the percentage of responses.

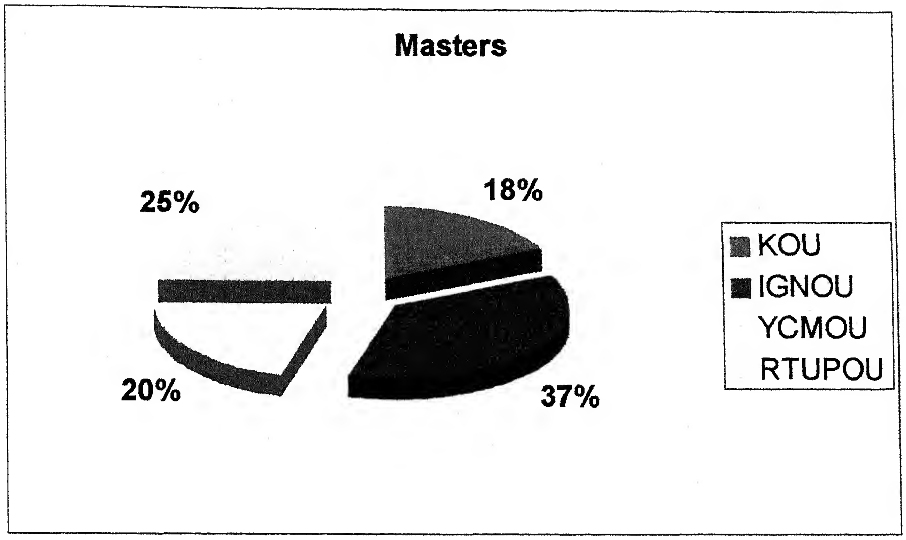
6.2 Category of Courses wise Distribution:

As a response In Kota Open University 09 certificate, 12 diploma, 21 bachelors and 08 students are perusing masters course. In Indira Gandhi National Open University 04 certificate, 06 diploma, 24 bachelors and 16 students are perusing masters course. In Y.C. Maharashtra Open University 05 certificate, 08 diploma, 28 bachelors and 09 students are perusing masters course. In U.P. Rajarshi Tandon Open University 06 certificate, 11 diploma, 22 bachelors and 11 students are perusing masters course.


Cirtificate
25%
37%
21%
17%
KOU
IGNOU
YCMOU
RTUPOU

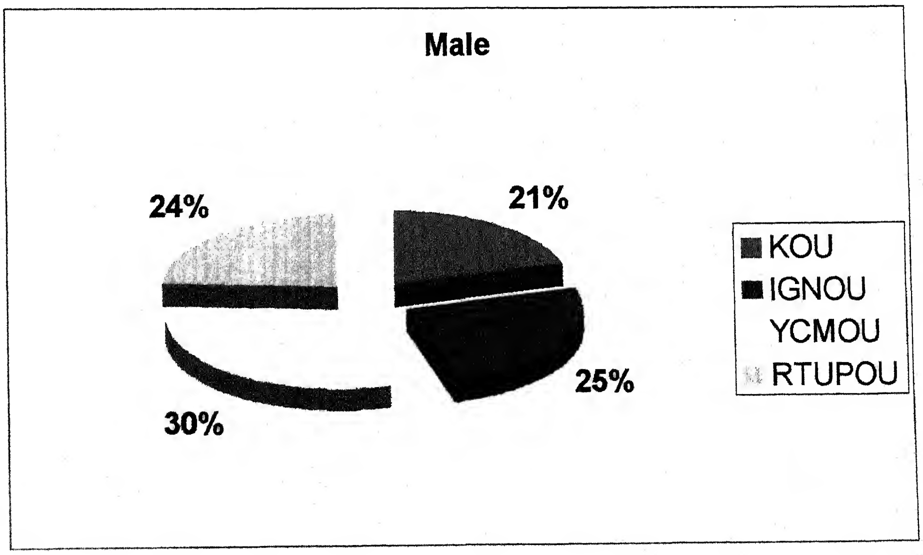

Diploma
30%
32%
22%
16%
KOU
IGNOU
YCMOU
RTUPOU

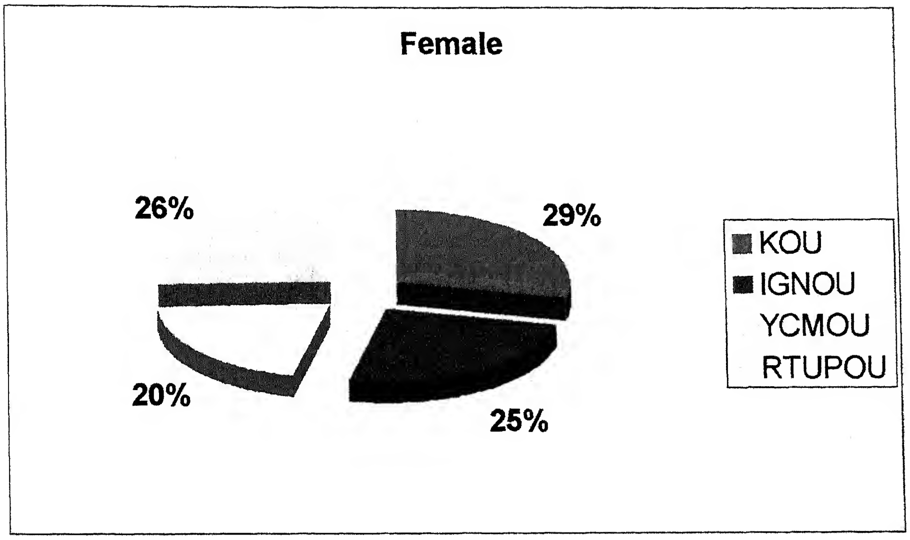

Bachelor
23%
22%
30%
25%
KOU
IGNOU
YCMOU
RTUPOU



6.3 Gender Wise Distribution:

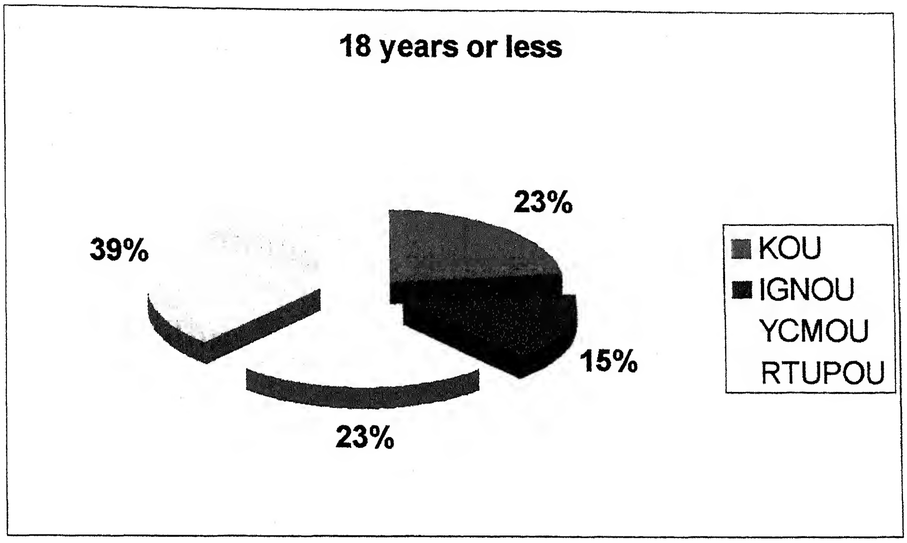
In terms of sex and specially male responses 19 responded in Kota Open University, 23 in Indira Gandhi National Open University 28 in Y.C. Maharashtra Open University and 22 male responded in U.P. Rajarshi Tandon Open University. At the same time 31 responded in Kota Open University, 27 in Indira Gandhi National Open University 22 in Y.C. Maharashtra Open University and 28 female responded in U.P. Rajarshi Tandon Open University.

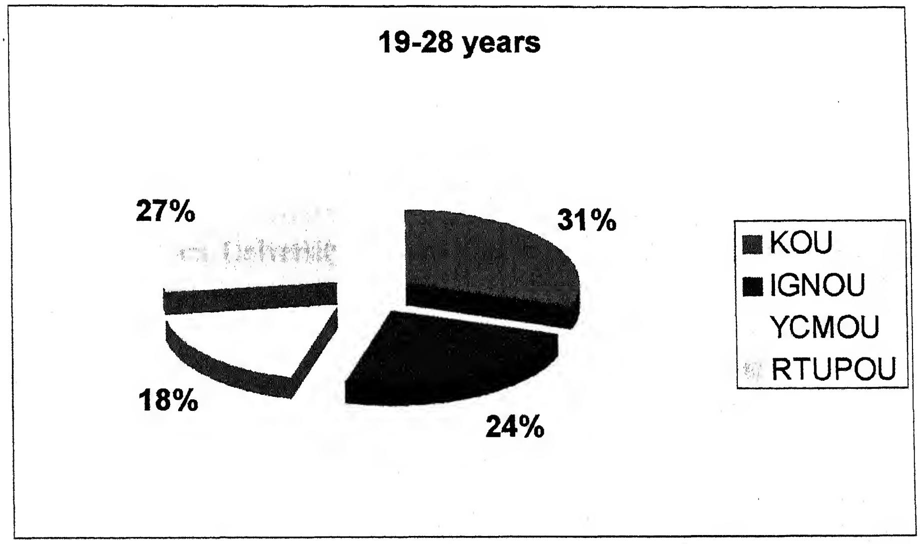


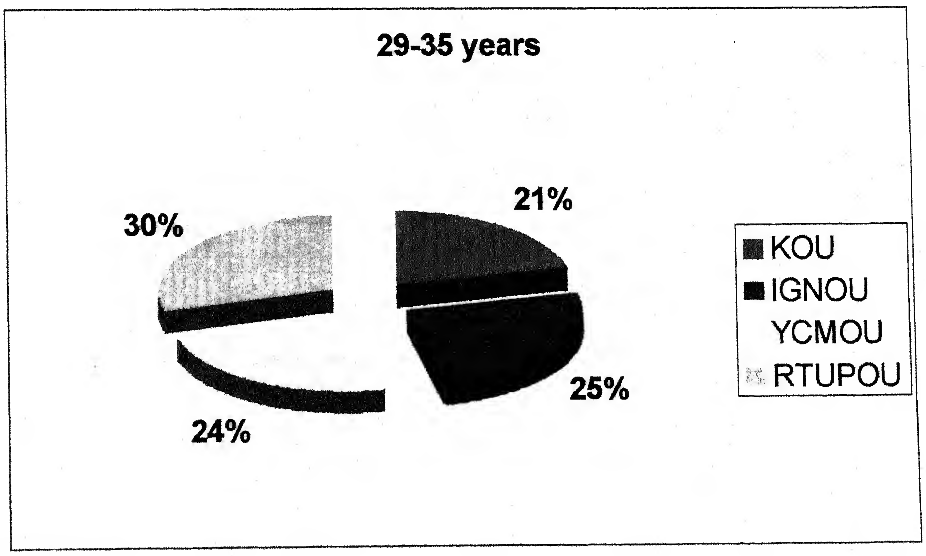


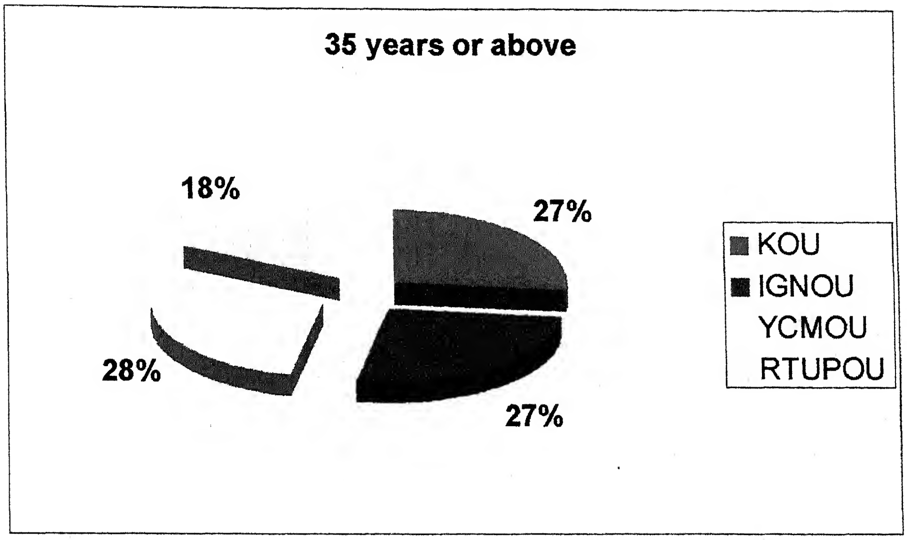
6.4 Age Wise Distribution:

As the age of 18 years or less 03 respondents are from Kota Open University, 02 from Indira Gandhi National Open University, 03 from Y.C. Maharashtra Open University and 05 from U.P. Rajarshi Tandon Open University. As the age of 18 to 28years 10 respondents are from Kota Open University, 08 from Indira Gandhi National Open University, 06 from Y.C. Maharashtra Open University and 09 respondents are from U.P. Rajarshi Tandon Open University. As the age of 29 to 35 years 16 respondents are from Kota Open University, 19 from Indira Gandhi National Open University, 18 from Y.C. Maharashtra Open University and 22 respondents are from U.P. Rajarshi Tandon Open University. As the age of 35 years or more 21 respondents are from Kota Open University, 21 from Indira Gandhi National Open University, 23 from Y.C. Maharashtra Open University and 14 respondents from U.P. Rajarshi Tandon Open University.


18 years or less
23%
39%
15%
23%
KOU
IGNOU
YCMOU
RTUPOU

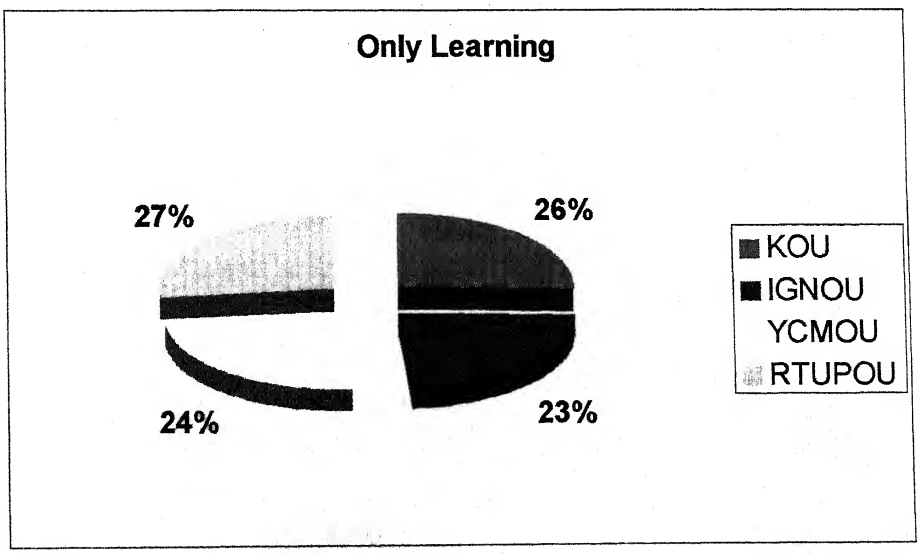

19-28 years
27%
31%
18%
24%
KOU
IGNOU
YCMOU
RTUPOU

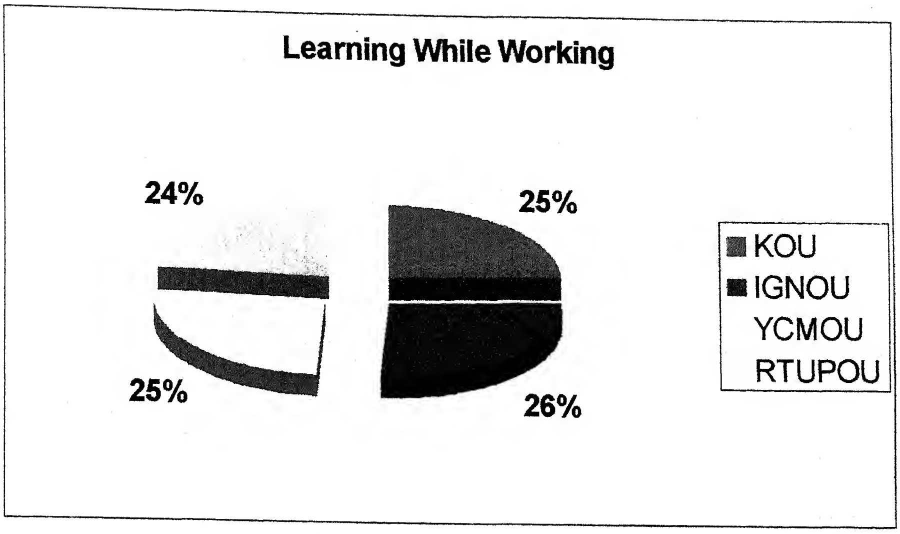

29-35 years
21%
30%
25%
24%
KOU
IGNOU
YCMOU
RTUPOU



6.5 Learning and Working Status:

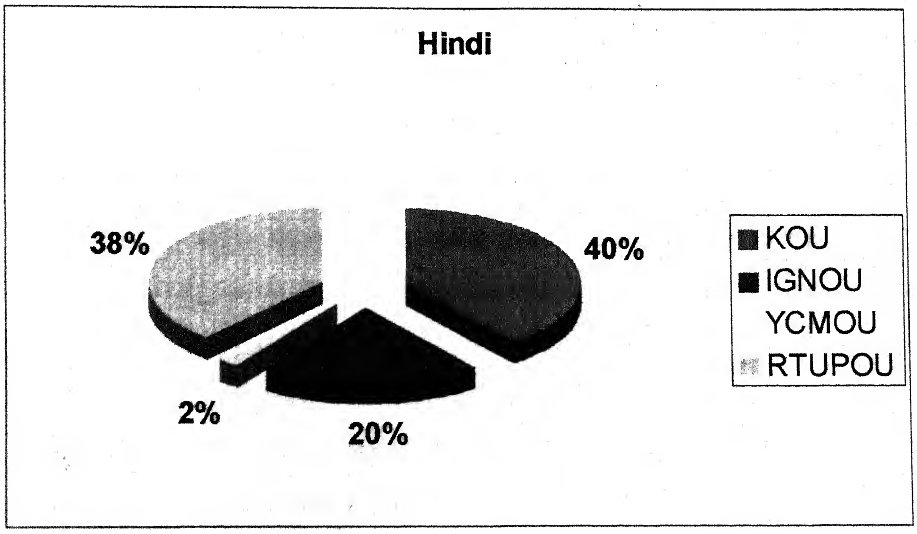
While checking the number of persons studying while working at the same time. 27 open learners in Kota Open University, 29 in Indira Gandhi National Open University, 28 in Y.C. Maharashtra Open University and 26 open learners in U.P. Rajarshi Tandon Open University are working at the same time while learning in various courses.

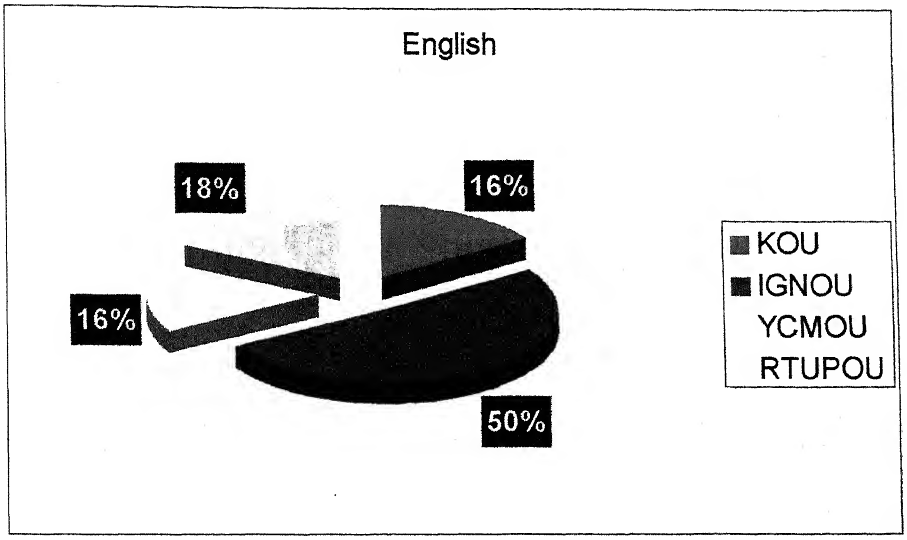


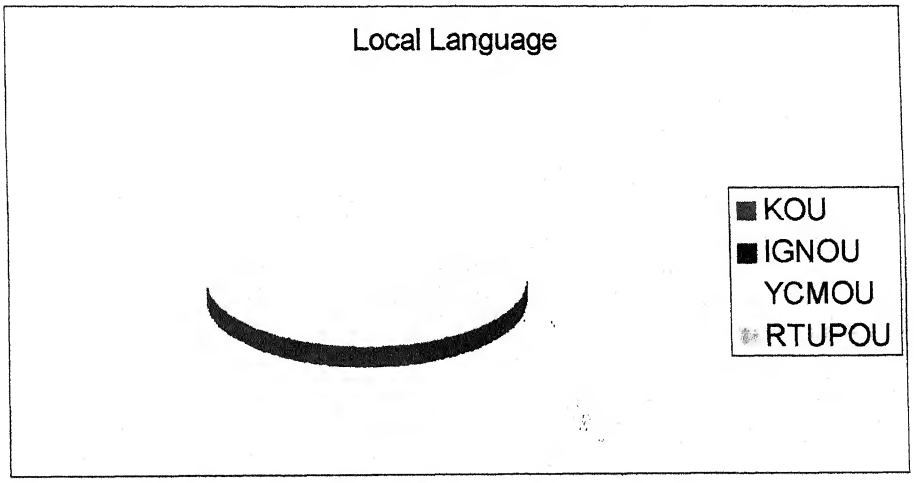


6.6 Language Wise Distribution:

While analyzing the instructional language wise distribution, it is found that as a medium of instruction hindi 41 respondents are in Kota Open University, 21 are in Indira Gandhi National Open University, 02 in Y.C. Maharashtra Open University and 40 are from U.P. Rajarshi Tandon Open University. As a medium of instruction English 09 learners are from Kota Open University, 29 are from Indira Gandhi National Open University, 09 are from Y.C. Maharashtra Open University and 10 learners are from U.P. Rajarshi Tandon Open University. While use of local language, only Y.C. Maharashtra Open University has 39 open learners opting local language as the medium of instruction.

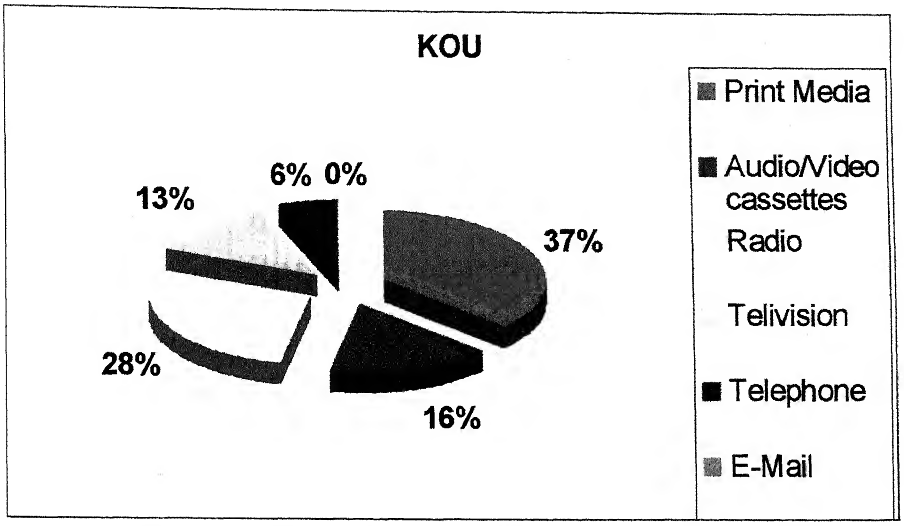


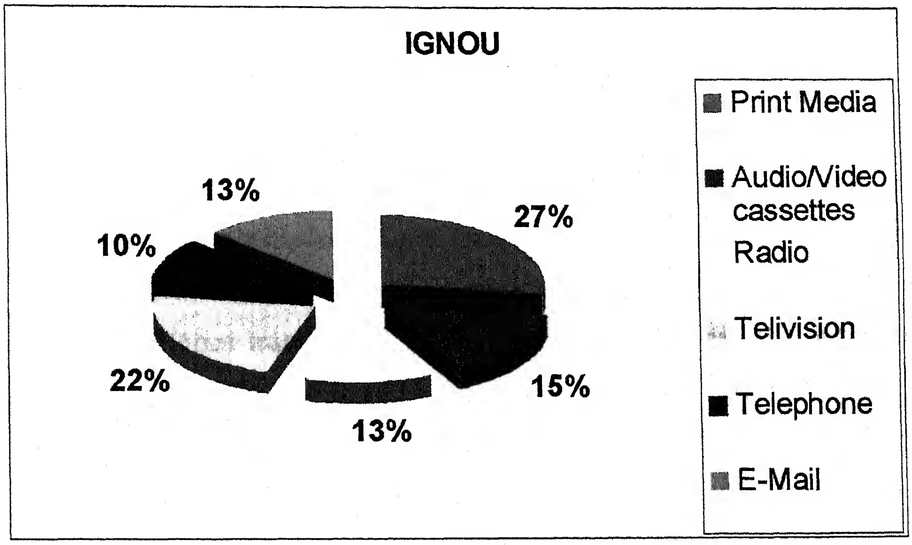


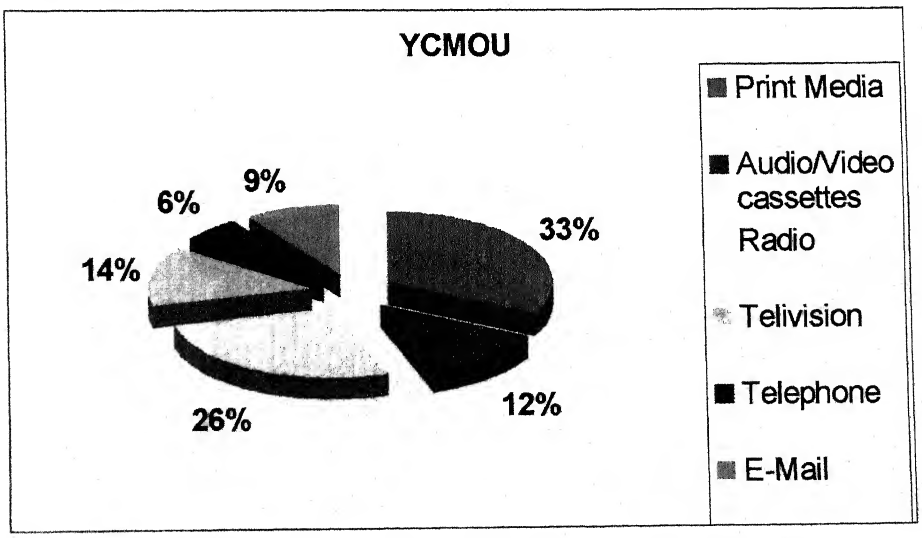


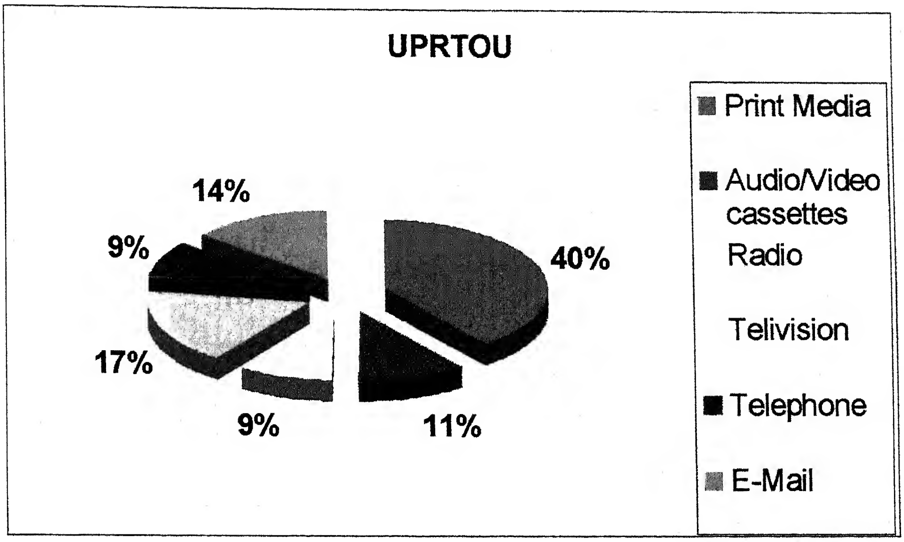
6.7 Technology/Medium Used for Instructions:

In Kota Open University 50 open learners are using print media, 22 are using audio/video cassettes, 38 are using radio, 18 are using television and 08 are using telephone. In Indira Gandhi National Open University 50 open learners are using print media, 28 are using audio/video cassettes, 23 are using radio, 41 are using television, 18 are using telephone and 24 open learners are using e-mail. In Y.C. Maharashtra Open University 50 open learners are using print media, 18 are using audio/video cassettes, 40 are using radio, 21 are using television, 09 are using telephone and 14 open learners are using e-mail. U.P. Rajarshi Tandon Open University 50 open learners are using print media, 14 are using audio/video cassettes, 12 are using radio, 22 are using television, 11 are using telephone and 18 open learners are using e-mail as the method of instruction and learning.


KOU
13%
6% 0%
37%
28%
16%
Print Media
Audio/Video cassettes
Radio
Telivision
Telephone
E-Mail

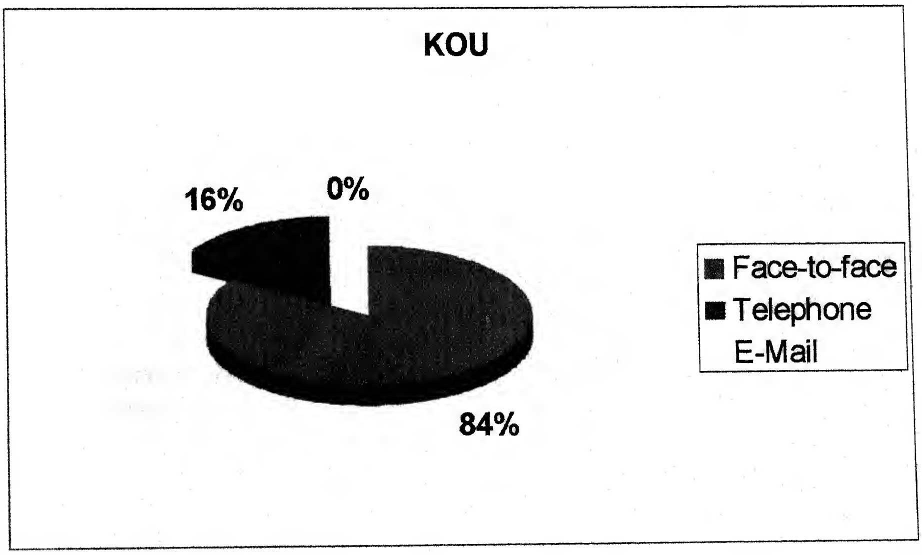

IGNOU
13%
27%
10%
22%
15%
13%
Print Media
Audio/Video cassettes
Radio
Telivision
Telephone
E-Mail

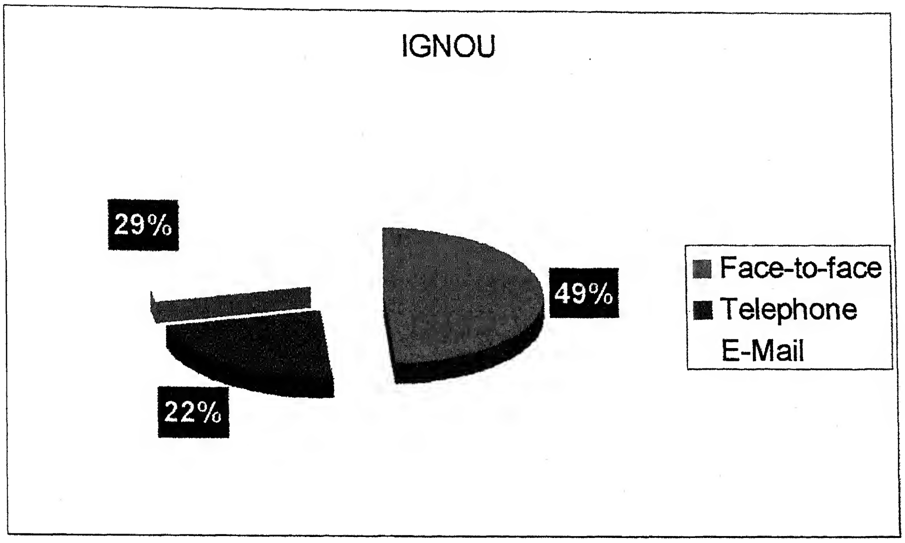

YCMOU
9%
6%
33%
14%
26%
12%
Print Media
Audio/Video cassettes
Radio
Telivision
Telephone
E-Mail

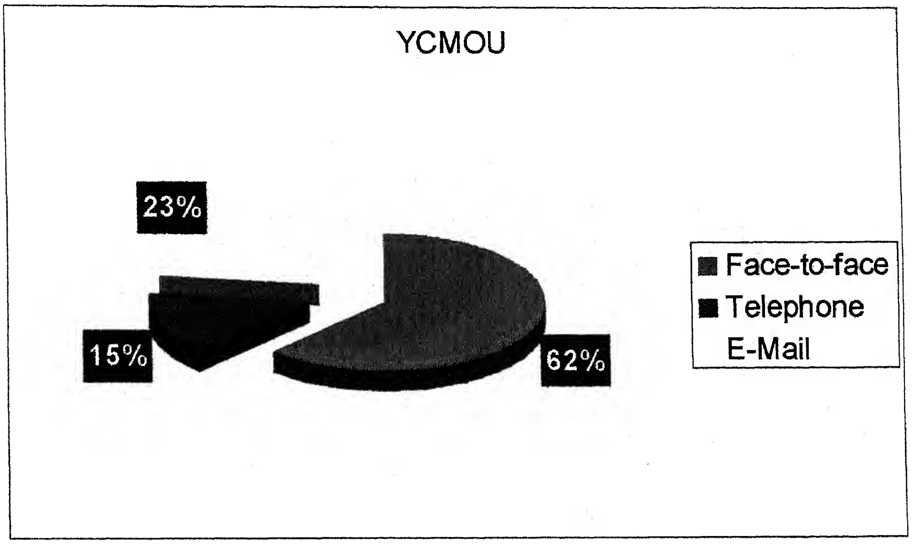


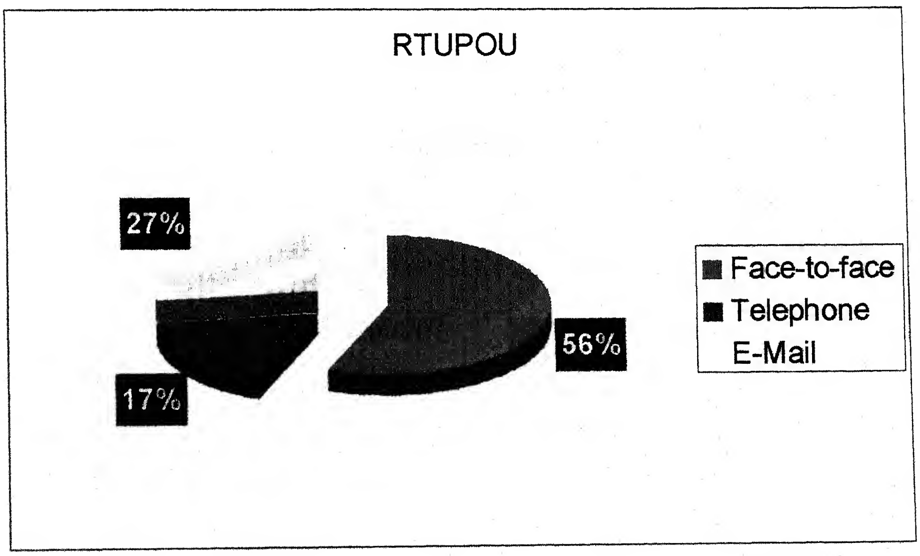
6.8 Communication Methods (Student-Teacher):

As a method of communication three methods are explored primarily and the response is in Kota open University 42 open learners prefer face to face communication, 08 telephone. In Indira Gandhi National Open University 40 open learners prefer face to face communication, 18 telephone and 24 prefer for e-mail. In Y.C. Maharashtra Open University 38 open learners prefer face to face communication, 09 telephone and 14 prefer e-mail. In U.P.R.T.O.U. 37 open learners prefer face to face communication, 11 telephone and 18 prefer e-mail as a mode of communication. Some other students also opt other methods, but not holding the priority position.



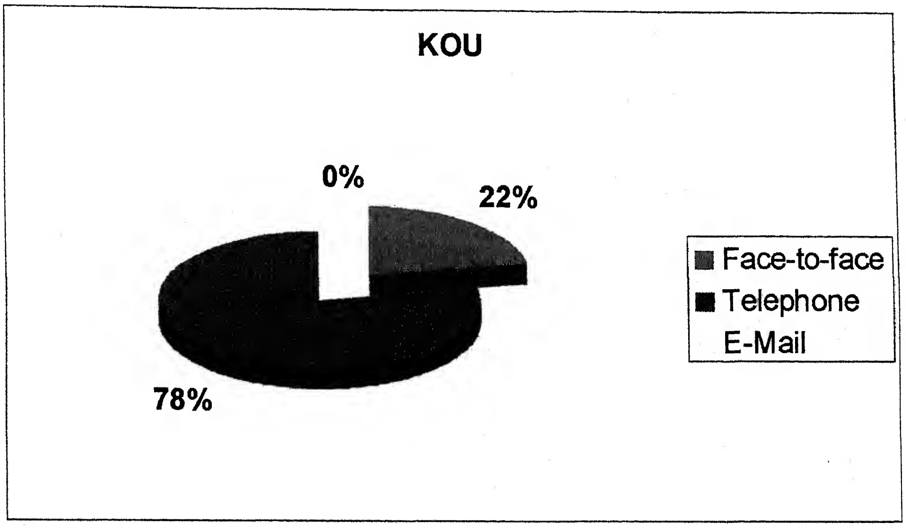

IGNOU
29%
49%
22%
Face-to-face
Telephone
E-Mail

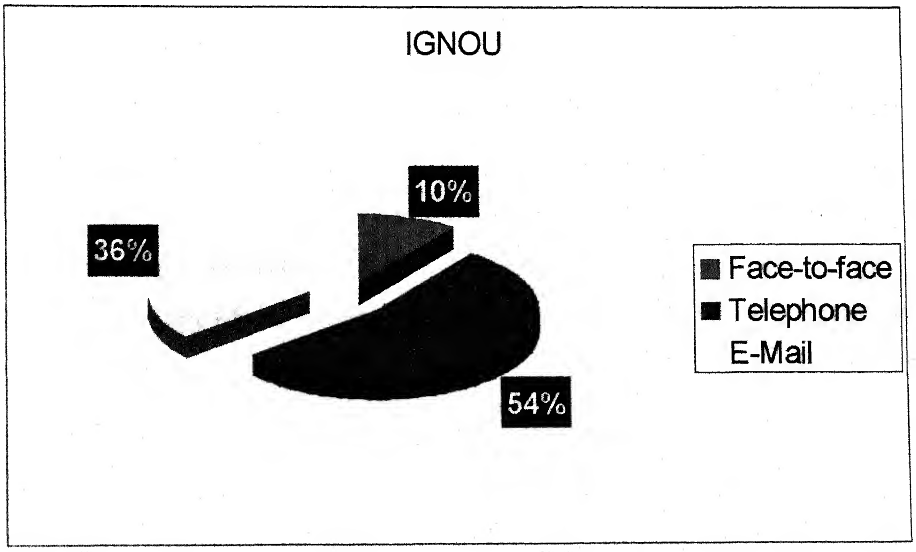

YCMOU
23%
15%
62%
Face-to-face
Telephone
E-Mail

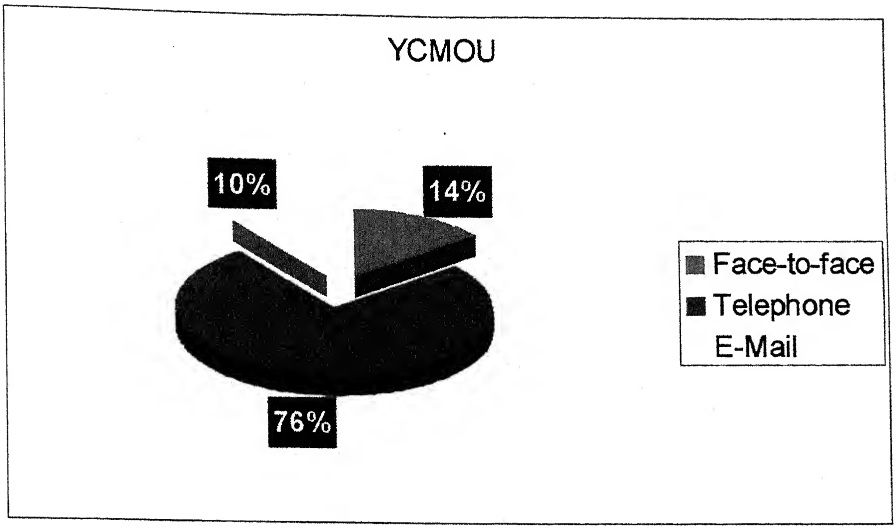

RTUPOU
27%
56%
17%
Face-to-face
Telephone
E-Mail

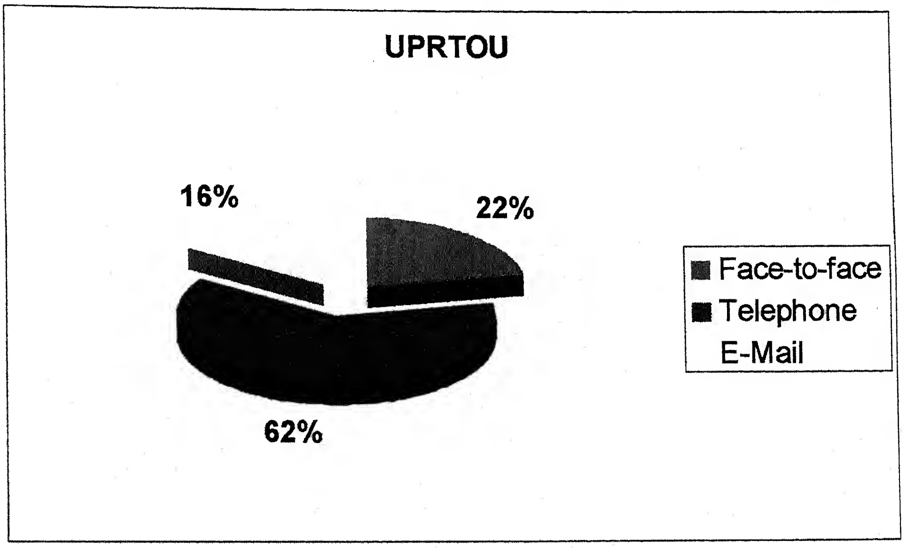
6.91 Communication Methods (Student-Student):

As a method of communication between students three mediums are primarily being used and the response is in Kota open University 11 open learners prefer face to face communication, 39 telephone. In Indira Gandhi National Open University 05 open learners prefer face to face communication, 27 telephone and 18 prefer for e-mail. In Y.C. Maharashtra Open University 07 open learners prefer face to face communication, 38 telephone and 05 prefer e-mail. In U.P. Rajarshi Tandon Open University 11 open learners prefer face to face communication, 31 telephone and 08 prefer e-mail as a mode of communication.



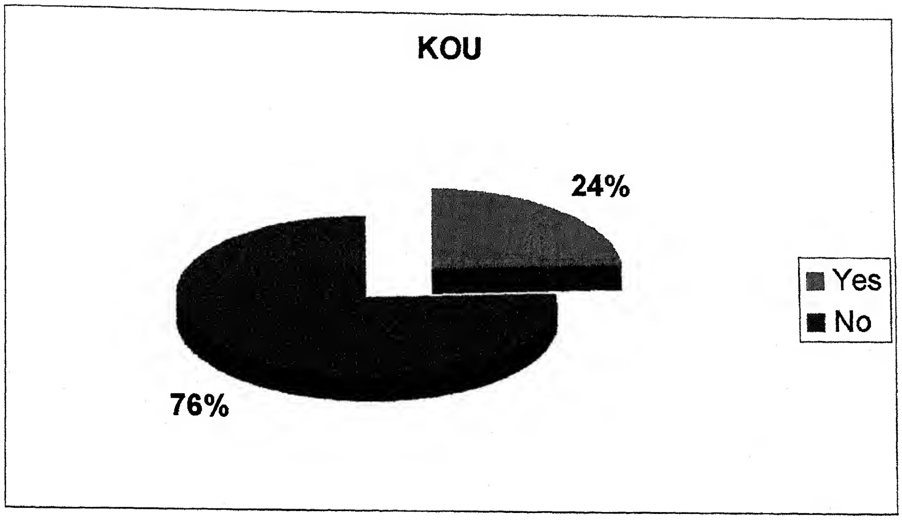


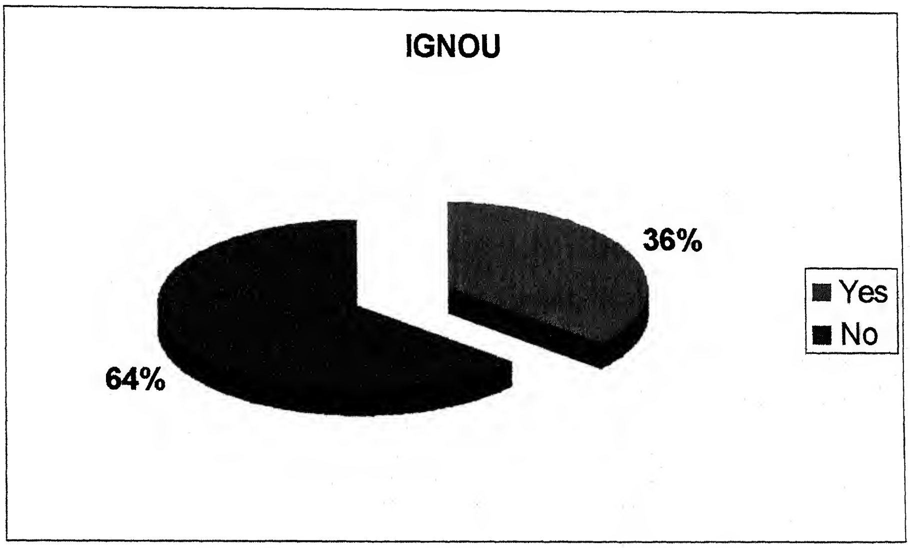


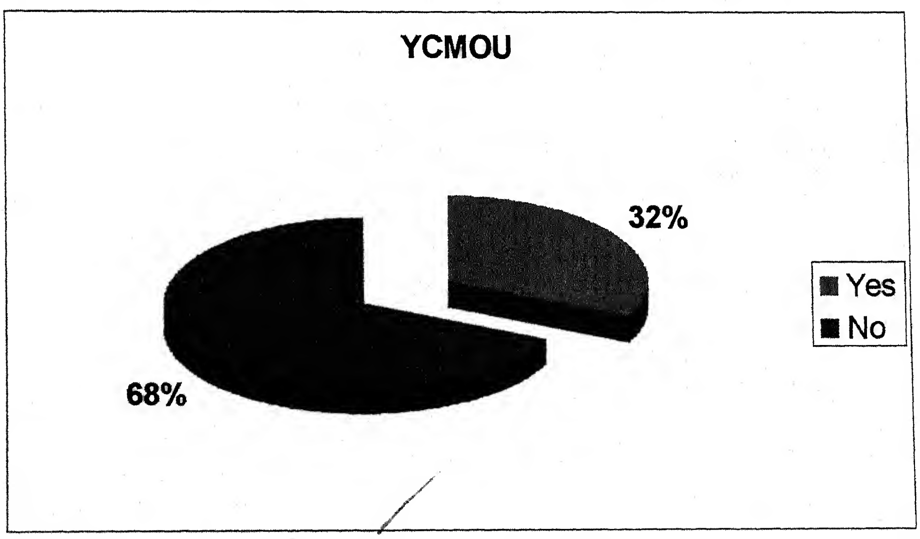


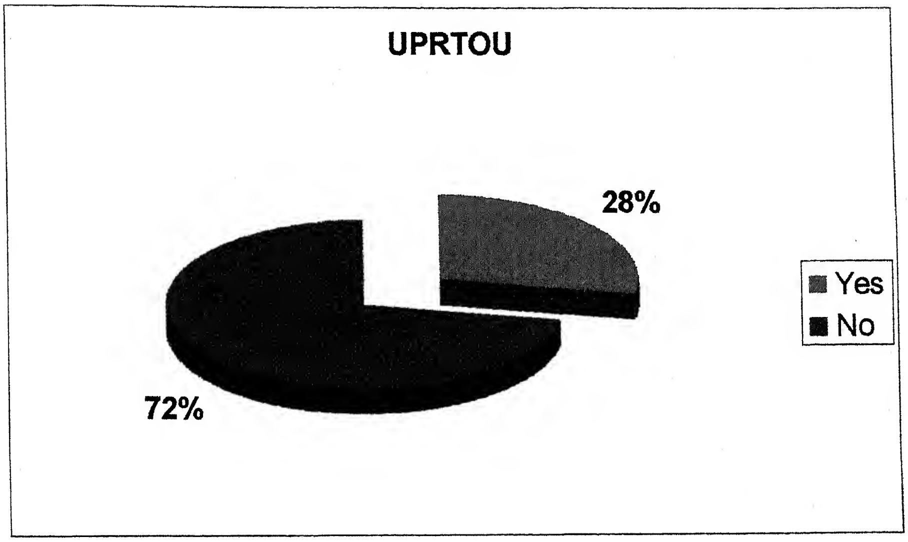
6.92 Use of University/Study Centre Library:

Data relating to the use of library whether it may be university library or library of a study centre shows its low usage. In Kota Open University only 12 persons, in Indira Gandhi National Open University 18, in Y.C. Maharashtra Open University 16 and in U.P. Rajarshi Tandon Open University only 14 persons are frequent user of the library.


KOU
24%
76%
Yes
No

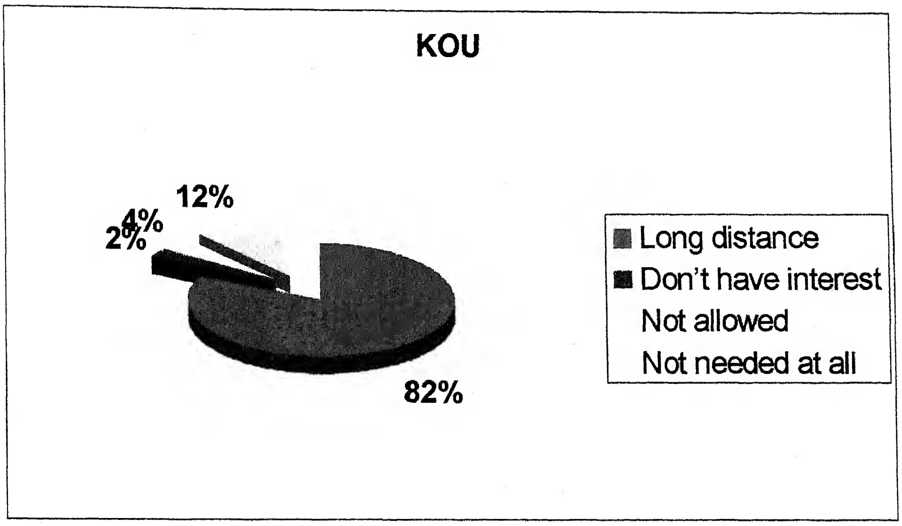

IGNOU
36%
64%
Yes
No

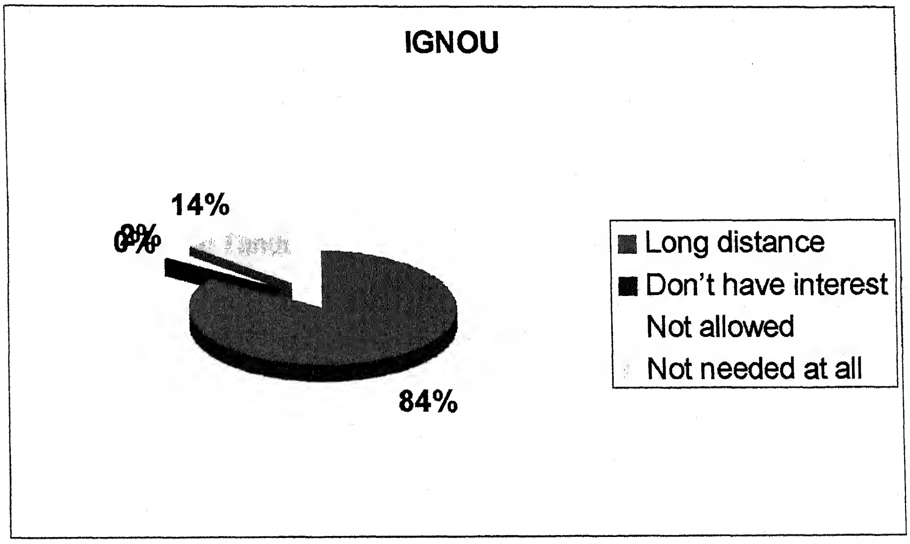

YCMOU
32%
68%
Yes
No

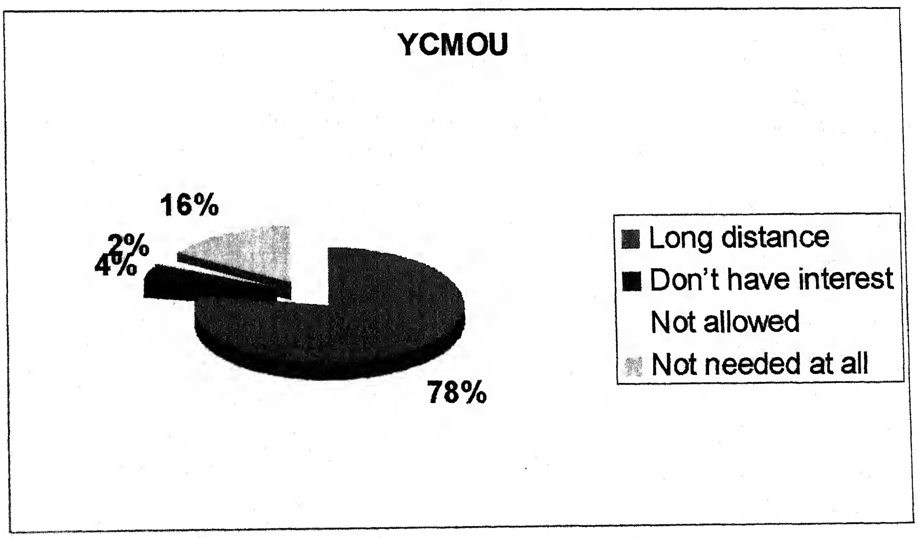


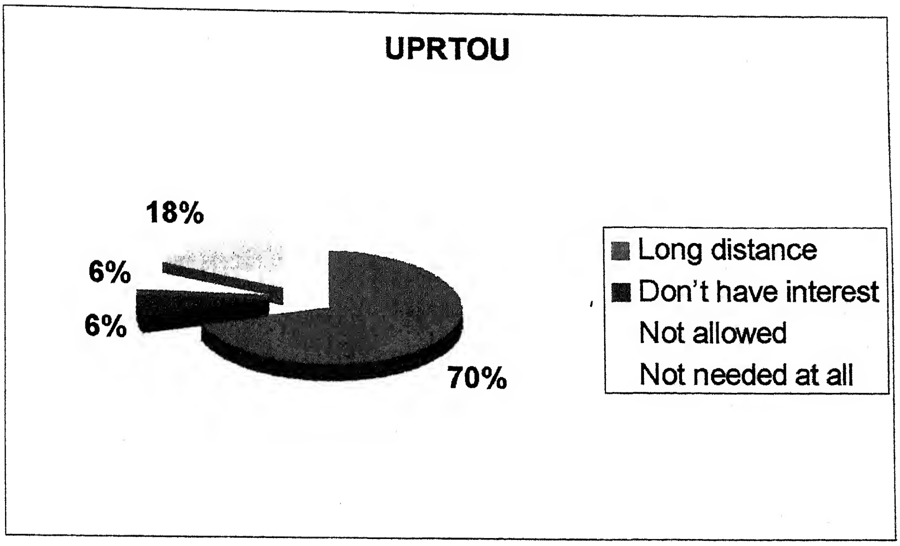
6.93 Reason of the Less Use of the Library:

The main reason behind low usage of the library specially of university library is long distance of the place where the library is situated. Besides this the response is, in Kota Open University 41 person says that it is long distant, 01 denies the facility to access library, 02 says that they are not allow to use library services and 06 person say that it is not required. In Indira Gandhi National Open University 42 person says that it is long distant, 01 says that they are not allow to use library services and 07 person say that it is not required. In Y.C. Maharashtra Open University 39 person says that it is long distant, 02 denies the facility to access library, 01 says that they are not allow to use library services and 08 person say that it is not required. In U.P. Rajarshi Tandon Open University 35 person says that it is long distant, 03 denies the facility to access library, 03 says that they are not allow to use library services and 09 person say that it is not required.


KOU
12%
4%
2%
82%
Long distance
Don't have interest
Not allowed
Not needed at all

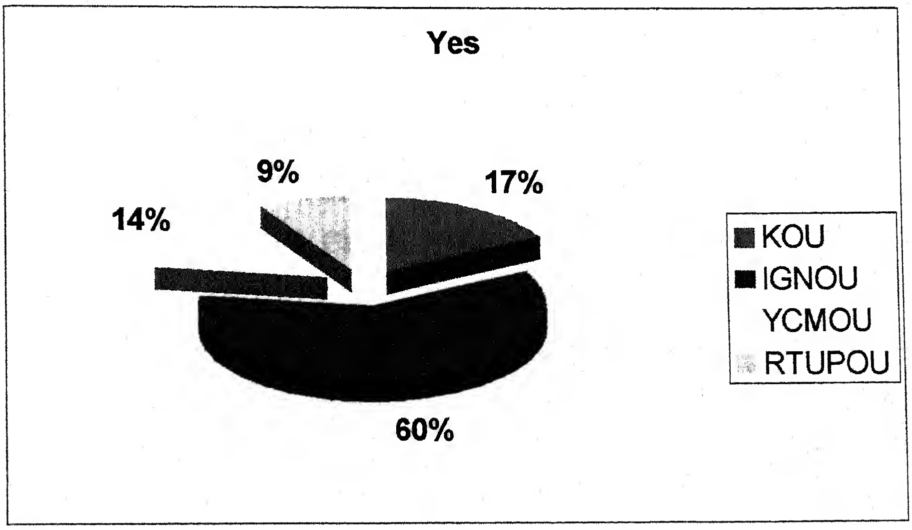

IGNOU
14%
84%
Long distance
Don't have interest
Not allowed
Not needed at all

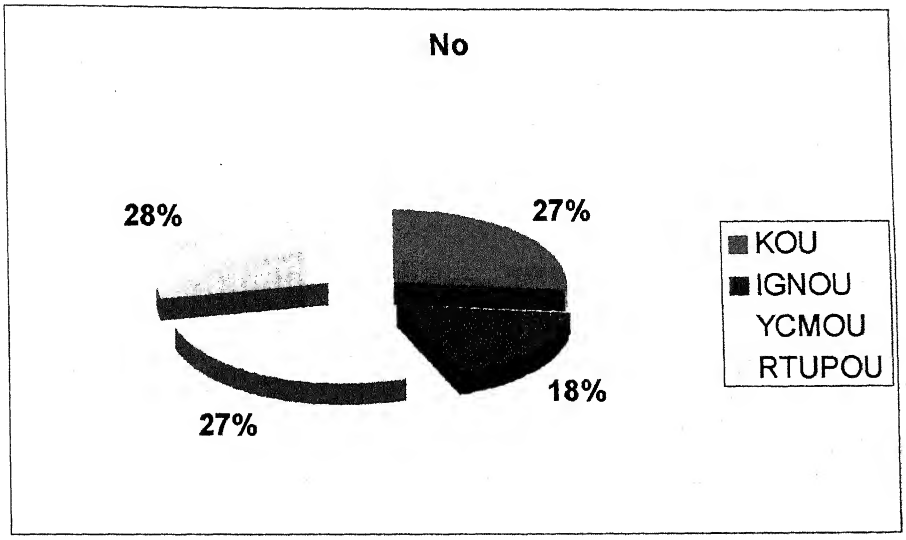

YCMOU
16%
2%
4%
78%
Long distance
Don't have interest
Not allowed
Not needed at all



6.94 Access of the Web Site of the Open University:

Number of students accessed university's web site is in Kota Open University 06, inIndira Gandhi National Open University 21, in Y.C. Maharashtra Open University 05 and in U.P. Rajarshi Tandon Open University 03.



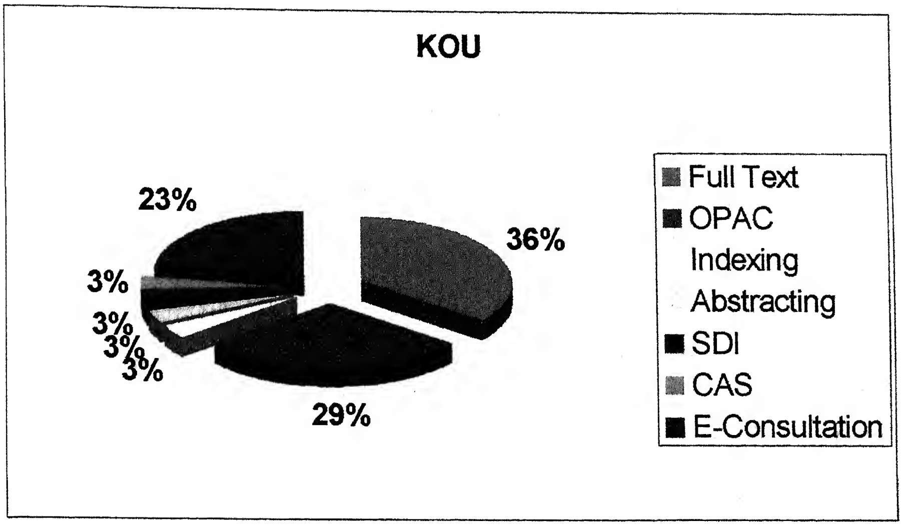


6.95 Library Access through Web:

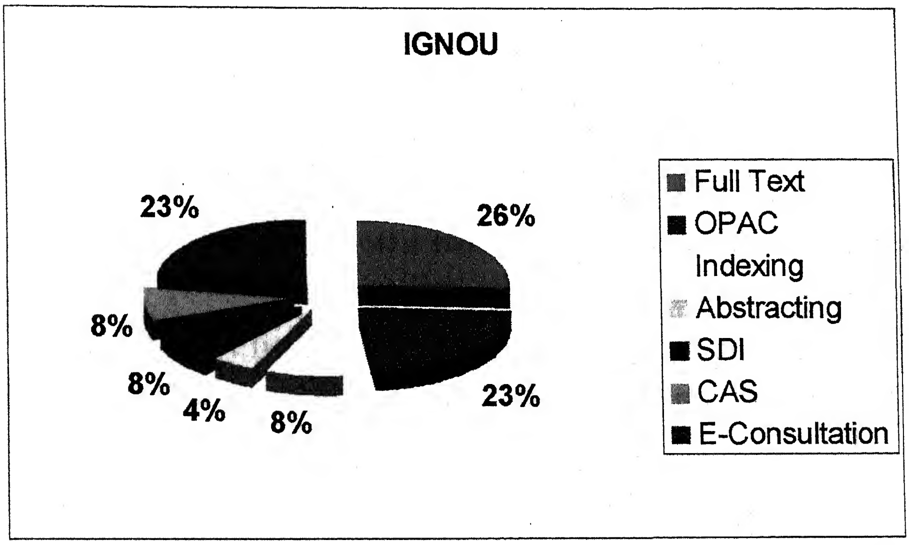
The count accessing university's web site is very low, and at the same time they are not aware about the web based library access. Indira Gandhi National Open University is an exception where 14 respondents have enquired about the library based services through web.

6.96 Need of Web Based Library Service:

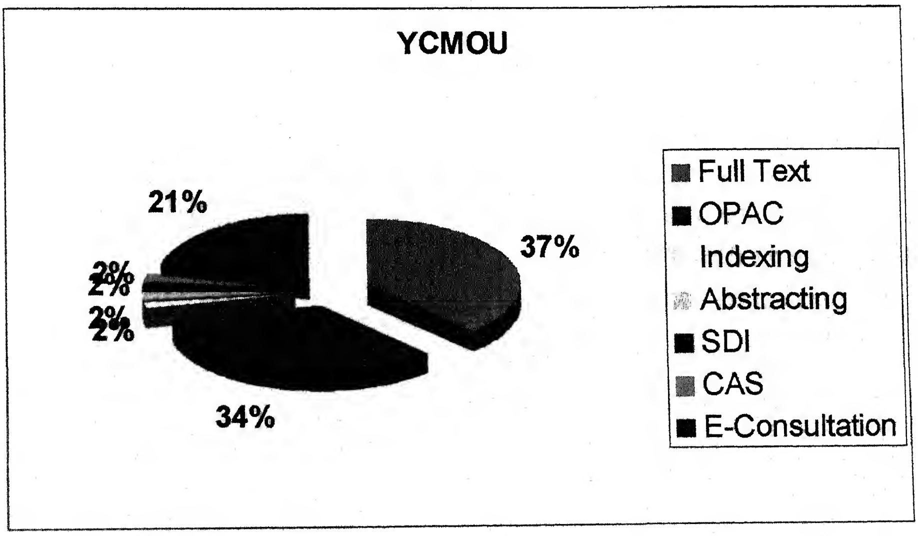
During the survey in Kota Open University 42 respondents required for full text service of library documents, 34 needs for OPAC, 04 demands for indexing, 04 want abstracting service 04 asks for SDI service, 04 want CAS service and 27 respondents want e-consultation. In Indira Gandhi National Open University 47 respondents required for full text service of library documents, 41 needs for OPAC, 15 demands for indexing, 08 want abstracting service 15 asks for SDI service, 15 want CAS service and 41 respondents want e-consultation. In Y.C. Maharashtra Open University 43 respondents required for full text service of library documents, 38 needs for OPAC, 02 demands for indexing, 02 want abstracting service 02 asks for SDI service, 02 want CAS service and 24 respondents want e-consultation. In U.P. Rajarshi Tandon Open University 41 respondents required for full text service of library documents, 36 needs for OPAC, 04 demands for indexing, 04 want abstracting service 04 asks for SDI service, 04 want CAS service and 29 respondents want e-consultation.



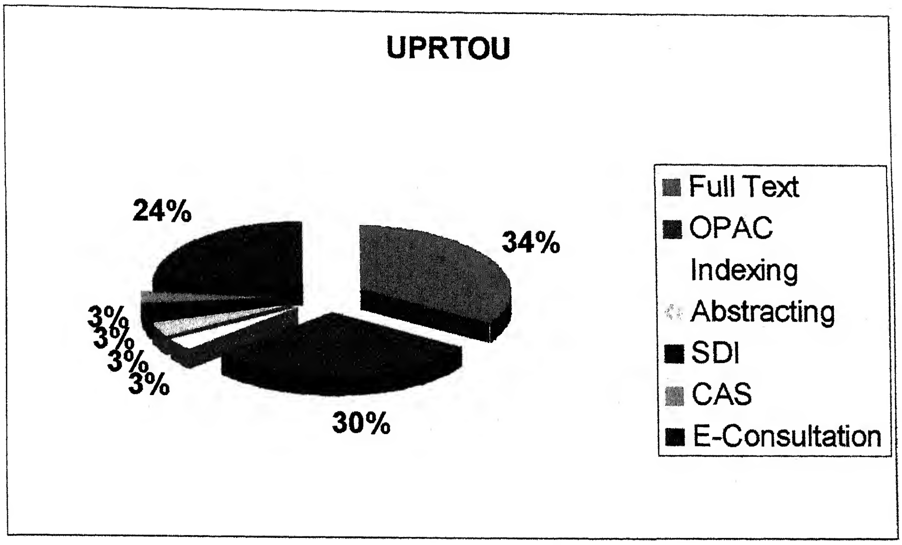
KOU
23%
36%
3%
3%
3%
3%
29%
Full Text
OPAC
Indexing
Abstracting
SDI
CAS
E-Consultation



IGNOU
23%
26%
8%
8%
4%
8%
23%
Full Text
OPAC
Indexing
Abstracting
SDI
CAS
E-Consultation

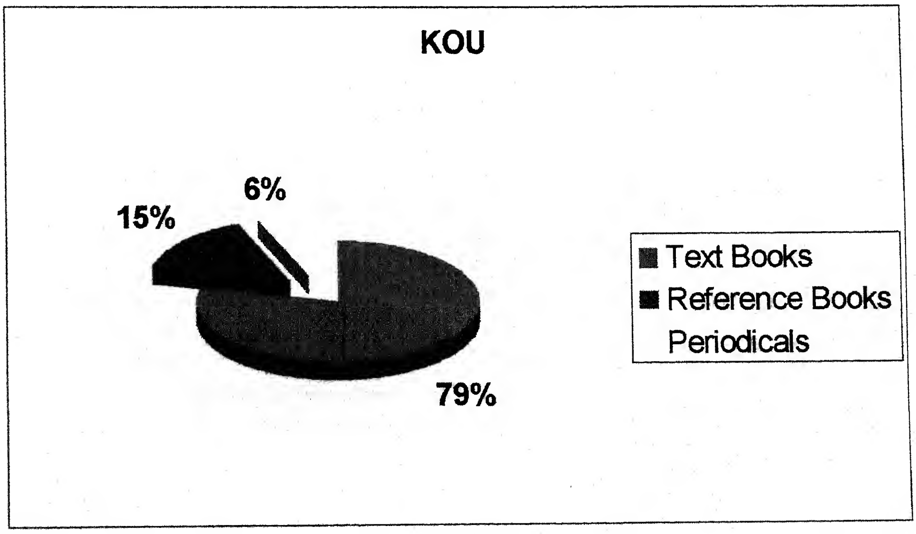


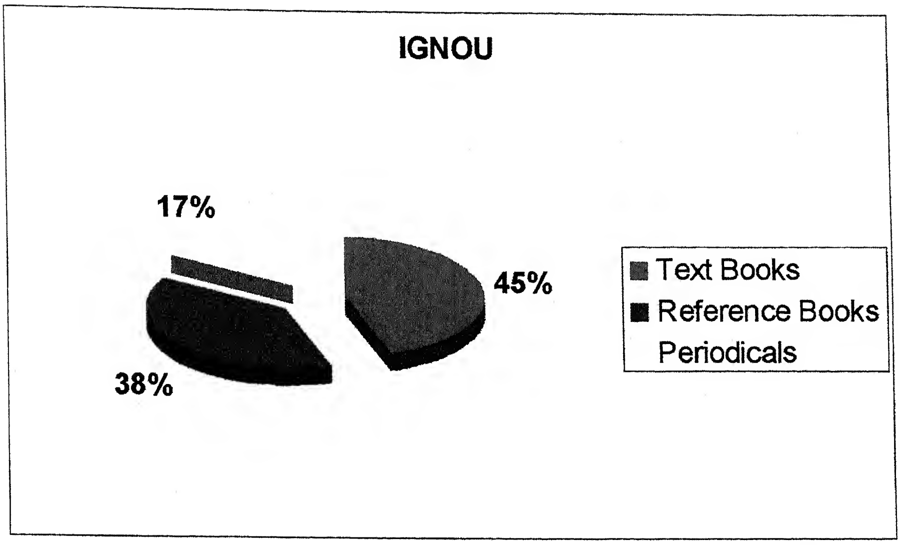
YCMOU
21%
37%
2%
2%
2%
2%
34%
Full Text
OPAC
Indexing
Abstracting
SDI
CAS
E-Consultation

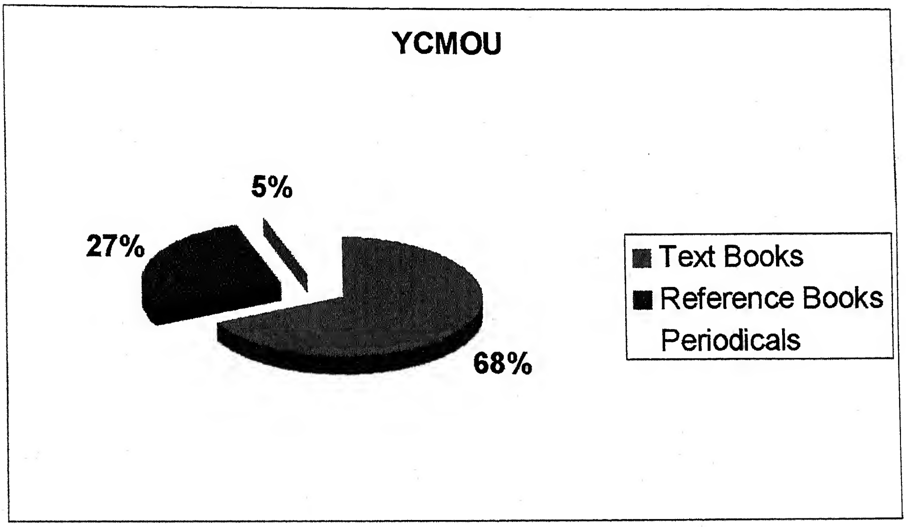


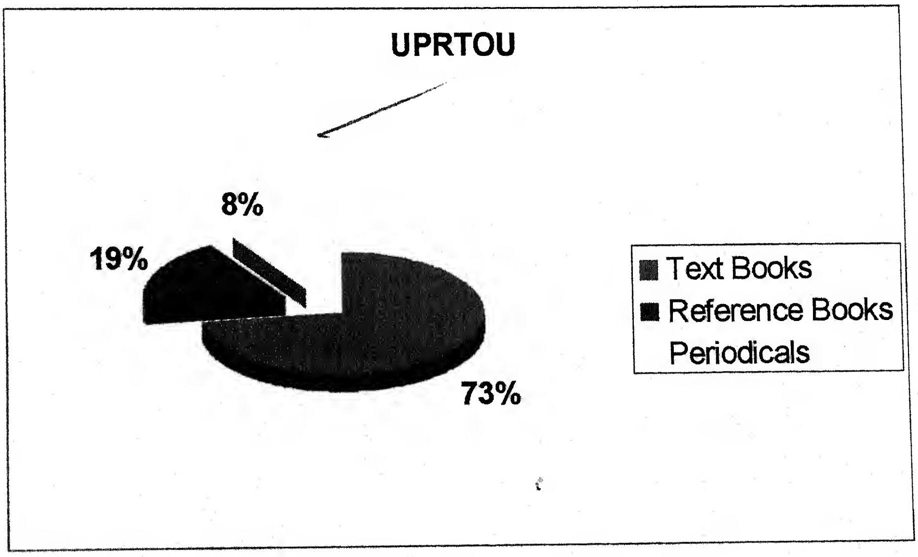
6.97 Type of Documents Accessible Through Web:

As the survey response 41 respondents want to access text books, 08 want reference books and 03 periodicals in Kota Open University, in Indira Gandhi National Open University 42 respondents want to access text books, 36 want reference books and 16 periodicals, in Y.C. Maharashtra Open University 40 respondents want to access text books, 16 want reference books and 03 periodicals, in U.P. Rajarshi Tandon Open University 43 respondents want to access text books, 11 want reference books and 05 periodicals.




IGNOU
17%
45%
38%
Text Books
Reference Books
Periodicals


YCMOU
5%
27%
68%
Text Books
Reference Books
Periodicals


UPRTOU
8%
19%
73%
Text Books
Reference Books
Periodicals

CHAPTER 7: FINDINGS AND CONCLUSION

Findings and conclusion:

After a long and intensive analysis it has been investigated that:

- To fulfill the basic objective of Open Learning System successfully it is essential to provide best services to Open Learners at their own specified and favorable conditions.

- Essentially provide consultancy at their own site without being physically present in the campus.

- Provide all material related to the study at learner's site.

- It is essential to provide all library services at Open Learner's site / place.

While checking/testing the hypothesis, the extensive analysis of data shows that

- it is possible to convert library material (documents) in digital format.

- It is possible to design a library's web page and to store digitized information on it, and

- the web based library's information can effectively and efficiently fulfill the library needs of a open learner.

While an extensive study of library and its services during the course of investigation it is found that all services related to the library fall under these categories:

- Document Delivery Service

- Reference Service

- Reprographic Services

- Documentation Services

It has been analyzed that it is not possible to fulfill our present objective until and unless the information professional changes the current infrastructure of a library to make it practically feasible to provide required services while using emerging tools and techniques for dissemination of information. Some of the proposed changes are:

- Acquisition of Digitised Information

- Computer based Processing of Library Material

- Digitization of existing Library Material

- Installation of ideal Web portal of the institution or library

- Make all Digital Information available through University / Library's Web portal

- Provide Web-based Consultancy related to the courses of study and day-to-day matters (On-line Consultancy).

- Check the feed back of users time to time while using ASP technique.

After an extensive research it has investigated that:

1. Most of the students are enrolled in graduate and postgraduate courses.

2. Students enrolled in certificate and diploma courses are mostly completing in professional subjects and are working at the same time.

3. Collected data shows that there is almost equal participation in the learning process by both of the genders.

4. Most of the open learners are from the age group of 35 years and above, this is also investigated that it is due to the specific promotional policies designed by the government for getting promotion benefits.

5. More that 50 percent of the open learners are learning while working at the same time.

6. Open learners from the states where regional language is the basic medium of communication prefer regional/local language.

7. English as a medium of instruction holds less acceptance but it is on the second place having second /optional priority.

8. Most popular medium of instructions among open universities is print media. Radio, television, audio/video cassettes are on the second place, depends on the facilities provided by the specific university.

9. Medium of interaction (student-teacher) is mainly face-to-face, while telephone is on the second position.

10. Most popular medium of interaction among students is telephone.

11. It is investigated that both the learners and the educators are very much curious to adopt new modes of instructions, i.e. E-Mail.

12. It is identified that the usage of library services among open learners is not satisfactory. The main reason investigated is the distance between the learner and the library.

13. Either the university has no web site or it does not provide enough learner based services. But it is investigated that the open learners are very much interested in the new technologies supporting the web-based services.

14. It is essential to provide full text information access of text books through web. The universities where research program are included, reference books and periodicals are equally important.

15. It is investigated to design a library web page having facilities to provide all interactive library services to support distant open learner not only to those having all capabilities but also to those who are suffering from the disabilities.

16. It is investigated to use latest web designing techniques to safe/barred the information from the hackers/unidentified users.

17. It is suggested that ISDN connectivity is better in terms of quality of services and the low expenditures.

18. It is essential to have a good knowledge of the source material.

19. All digitized information should be stored in a centralized digital master file, which must be saved in duplicate on several media.

20. For better resolution it is suggested to capture the photographs with an appropriate Dots Per Inch(DPI), Binary and Bit Depth values.

21. It is suggested to use flatbed scanner for photographic material or transparencies.

22. It is suggested to use digital camera or overhead scanners for bound volumes or over sized flat materials.

23. While choosing scanning equipment it is suggested to check the actual requirements and match them with the capacity of the scanning equipment.

24. While choosing scanning software it is suggested to select a S/W with an ability to run batch scans.

25. Before starting the digitization work it is suggested that a healthy understanding with the copyright owner must be taken place first.

26. While running a big digitization unit, it is suggested to inspect the scanning and the post scanning process and products after a specified duration periodically.

27. It is suggested that after scanning the original document save it in the appropriate format, it may be PDF, GIF, JPEG or any other.

28. It is suggested to use TIFF format for optimization and manipulation purpose.

29. It is investigated that JPEG is a good file format for web-based services because of its widely supported by web browsers.

30. It is suggested that MrSID should be used to store large format cartographic materials.

31. It is suggested that DjVu is a particularly good file format for storing handwritten letters, manuscripts and early printed materials.

32. It is investigated that the Height: 12.65 in., Width: 8.18 in., Brightness: 125 (approx.), Contrast: 130 (approx.), Image Type: Black and White Photo, File Type: JPEG, Image Quality: 600 dpi must be followed for better and uniform results.

33. Before mounting the scanned textual information on to the web it is suggested to convert it in a searchable database using appropriate metadata.

LIMITATIONS AND FURTHER AREA OF RESEARCH:

The present study is confined to "An Investigation on Web Enabling of Library Material in Open Learning System", and thus limit its scope to the library-based services provided by open universities only. And thus it is a suggestion to carry on the similar type of work for all other universities, educational institutions and research organizations, as the information seeking behavior of library users is almost similar. Besides this, with meager infrastructure facilities it was a difficult task of conducting research in such a specialized technical area, which is growing rapidly. Thus it is the responsibility of government and other funding agencies to further support this as an important area of research.

BIBLIOGRAPHY

1. AGARWAL (S N), KHAN (R R) and SATYANARAYANA (N R)*Ed.* Perspectives in library and information science (Vo 1 – 2) Print house, Lucknow. 1982.

2. AIKEN (SG); DATTWITZ (MJ); MCJANNE (CL) and CONSAUT (LL). "Biodiversity among Festuca(poaceae) in North America: Diagnostic evidence from DELTA and clustering programs, and an INTKEY package for interactive illustrated identification retrieview", Canadian Journal of Botany 75(9) p1527-1555 1997

3. ANAND (S). University without walls: The Indian perspective in Correspondence Education. New Delhi: Vikash
Publishing House, 1979.

4. BANERJEE (BN). The role of the University library in teaching and research. *IASLIC Bulletin*. 8; September 1963.

5. BARMELY (Gerald). A history of library education. Clive Bingley, London. 1969.

6. BATES (AW). "Third Generation Distance Education: The challange of new Technology", Research in Distance Education 3(2) pl0-16 1991

7. BESEMER (H) and Veerman (I). "Agricultural information on the Internet: What is out their and how to find it ", Quarterly Bulletin of IAALD 40 (2-3) p 60-67, 1995

8. BLAKE (EM). The Library College movement. Drexel Lib. Q.7,1971.

9. BOARAH (Swapana) *ed.*, Distance education. Delhi, Amar Prakashan, 1987.

10. BOJER (OQ); MURALI (NS); SECHER (BJM); BOSSEN; KURE (H) and KRISTENSEN (AR) *Ed.* "Sortinfo: an information and decision support system for variety selection", Proceedings of the First european Conference for information Technology ib Agriculture, Copenhagen, Denmark, 15-18 June 1997, p235-238 1997

11. BONN (George S), *Ed.* Library education and training in developing countries. East – west center press, Honolulu. 1967.

12. BOWEN (JP). "The World Wide Web Virtual liabrary of Museums", Information services & use 15 p317-324 1995

13. BRAMLEY (Gerald). World trends in library education. Clive Bingley, London. 1975.

14. BROWN (S). "The European Collaborative Backcross Database MBz is now available on the World Wide Web" Mammalian Genome 6(4) p223 1995

15. CARNOVSKY (Leon). Evaluation and accreditation of library scholars. *Library Quarterly*. 37, 4; Oct 1997; 334 – 344.

16. CHIB (SS). Teaching by correspondence in India. Light and life publications, New Delhi. 1977.

17. CLARKE (J). Resource based learning for higher and continuing education. Croom Helm, London. 1982.

18. CLOYES (Key). "The journal from vision to reality of a virtual library", Specialliabraries 85(4) p253-257 Fall 1994

19. CONTE (E). "Database on pesticides accessed through the internet", Proceedings of

the first workshop of the European Network for information Technology in Agriculture, Rome, Italy, 28-29 Nov 1996 p155-162 1997

20. DALE (Sheila). Adult independent learning project: Work with adult self directed learners in public libraries. Jou. of Librarianship. 11(2), April, 1979.

21. DASGUPTA (RK). New education policy and librarians. Herald of Library science. 25 (1-2) 1986. P 46-56.

22. DATT (Ruddar). Planning and development of distance education. *Journal of higher education.* 9, 3; 1984; 305 – 307.

23. DATT, (Ruddar). Role of distance education. University News. 24(8),1984. p5-8

24. DAVIS (GL); GILMAN (E.F) and DECK (HW). "An electronically based horticultural information retrieview system", HortTechnology 6(4) p332-336 1996

25. DESHMUKH (K G). Distance education and its future through communication technologies. *University News.* 32, 47; 1985; 9-13.

26. DESHMUKH (K G). Genesis and the growth of distance education. University News.24(42), 1986. PP. 3-6.

27. DEWAY (J). Democracy and education. Macmillan, New York, 1916.

28. ERDOS (R.F). "Teaching by correspondense: Unesco source book'" London: Longman, 1967

29. ERDOS (RF). Teaching by correspondence. Longman, London. 1967.

30. FAIBISOFF (SG) and WILLS (DJ), Distance education definition and overview. Journal of Education for Library & Information Science. 27(4), 1987.

31. FAURE (E)…et.al. Learning to be: The world of today and tomorrow. UNESCO, Paris. 1970.

32. FISHER (K), Library Services for adult continuing education and independent learning: Guide. London, Library Assn., 1988.

33. FISHER (K), Library Services for University adult education. 53, 1981. PP. 372 - 375.

34. FOWELL (S) and LEVY (P). Developing a new professional practice: A model for networked learner support in higher education. Journal of Documentation. 51, 3; 1995; 271– 280.

35. GEORGE (KM) *ed.*, Indian Libraries: Trends and perspectives. Calcutta, Orient Longman, 1985.

36. GOLDEN (BL), "The World Wide Web: HTML, CGI and Java application in animal Breeding", Proceedings of 6th world congress on Genetics Applied to live stocks production, Armidale, Australia, Jan 11 - 16, 1998 p409-414 1998

37. GUPTA (Pawan) and Usha, Education Policy and public libraries.ILA Bulletin. 21 (3-4), 1985-96 PP. 111 - 117.

38. HARBERS (GW); HOOGENBOON (G), " The use of real-time database applications to develop dynamic web content for information delivery in time of crisis", 2000 ASAE- Annual international meeting, Milwaualle, USA, 9-12 Jul 2000 p 1-10 2000

39. HARGER (C) et.al., "The Genome Sequence Database (GSDB) improving data quality and data access", Nucleic Acids Research 26(1) p21-26 1998

40. HARGER (C) et.al., "The Genome Sequence Database version 1.0 (GSDB): from the low pass sequence to complete genomes", Nuclic Acids Research 25(1) p18-23 1997

41. HAWES (ME), The role of the large public library in adult education. Library Trends.8(1), July 1959.

42. HENDRIK (Vandes Zee), The Public library as an opening centre. Journal of librarianship. 20 (1), Jan, 1988 PP. 1-15

43. HENEGHAN (PA) et.al., "Pond FX: An internet Tool for exploring invertebrate population recovery", Aspects of Applied Biology 53 p263-270 1999

44. HIGHAM (Norman). The library in the university: observations on a service. Andre Deutsch, London. 1980.

45. HOLMBERG (B). Correspondence education. Malmo, Hermods.1967.

46. HOLMBERG (B). Status and trends of distance education. Kogan, London. 1981.

47. HOLMBERG (Borjee). Tutoring distant students. *Epistolodidaktika*.1; 1977;65-74.

48. HULMBERG (Borjee). Distance education. Kegan Paul, London. 1977.

49. HUTCHINSON (B); HENZEL (J); RAM (S) and TIMMERMANN (B). "Managing information products for the Bioactive Agents from Dry land Biodiversity of Latin America Project", Quarterly BuHetin pf IAALD 44(3-4) P 163-172 1999

50. INDIA, Ministry of Human Resource and Development. National Policy on education. MHROD, New Delhi. 1986.

51. INDIA. Education commission (1964-66) (chairman: D.S. Kothari) Educational planning and the national policy. NCERT, Delhi. 1971.

52. INDIA. Government of India, "Report of III Planning Commission, 1960-1965" New Delhi

53. INDIA. Government of India, "Report of the Education Commission, .19641966", New Delhi, 1966

54. INDIA. Indira Gandhi National Open University. Growth and Phi losophy of distance education in distacne education course

55. INDIA. Indira Gandhi National Open University: A Profile. IGNOU. New Delhi. 2001.

56. INDIA. Report. Education commission (1964 – 66), Ministry of Education, Govt. of India, New Delhi. 1966.

57. INDIA. U.G.C. (India). Development of Library facilities in Universities and colleges.U.G.C. New Delhi.1968.

58. INDIA. U.G.C. (U.K.) Committee on libraries (1963) Chairman: Thomas Parry) Report, HMSO, London. 1967.

59. INDIA. U.G.C. Committee on University and college libraries, (1957) (Chairman: S.R. Ranganathan). Report. U.G.C. New Delhi. 1965.

60. ISAAC (K A) and DEVARAJAN (G), Ed. Libraries in distance education. Ess Ess publications, New Delhi. 1989.

61. JAGANATHAN (Neela). Collection development at the libraries teaching universities with special reference to the Indian situation. ILA Bulletin.22(3) Oct. - Dec., 1987. PP.140 - 148.

62. JEEVAN (VKJ). Distance or distress education for librarianship. Herald of Library Science. 37, 3 – 4; 1998; 196 – 204.

63. KANJILAL (Uma). Education and training of library and information science professional through distance mode: Challenges for the Indira Gandhi National Open University in the next millennium. In , Ed. Channels of learning and delivery in Open University system. IGNOU, New Delhi. 1997.

64. KAUL (RN) *et. al., eds* Studies in distance education. New Delhi, IGNOU, 1988.

65. KAULA (PN) Higher education and libraries. Herald of Lib.Sc.25 (1-2), 1986. PP. 46 - 56.

66. KAYE (Anthony) and RUMBLE(Greville), Ed. Distance teaching for higher and adult education. The Open University press, London. 1918.

67. KAYE (E). Distance education. International Encyclopedia of Education: Research and studies. V3. Perganon, New York. 1985.

68. KEEGAN (D) and HOLMBERG (G) *Eds.* Distance education: Internal perspectives.Croom Helm, London.1983.

69. KEEGAN (D). Founctions of Distance Education. London, Croom Helm, 1986.

70. KENT (Allen) and LANCUR (Harold), *Ed.* Encyclopedia of Library and Information science. Marcel Decker, New York. 1972.

71. KHAN (Inayat). "Teaching at a distance", Delhi: Amar Prakash_ 1989

72. KHAN (Inayat). Distance teaching. Amar prakashan, Delhi.

73. KHURSHID (Anis), *Ed.* Library education in India, Pakistan and Iran. India Bibliographies Bureau, New Delhi. 1987.

74. KOLASKAR (AS) and NAIK (PS). "Online identification of viruses" Journal of Microbiology, Immunology and infection 33(2) p69-78 2000

75. KOUL (B N); SINGH (B) and ANSARI (MM), *Ed.* Studies in distance education. Association of Indian University, New Delhi. 1988.

76. LAFONTAINE (DA); MERCURE (S); PASSION (V) and PARREAULT (JP). "The viriod liked RNA database", Nucleic Acids Research 26(1) p190-191 1998

77. LENGRAND (P). An introduction to life long education. UNESCO, Paris. 1970.

78. LEX, J. "Virtual agriculture: Agriculture in the information in the information society", Zeitschrift fur Agrainfomatik 3(6) p125-130 1995

79. LIN (Nan) et.al. "Decision information systems for citrus: software implementation and testing", Proceedings of the Florida State Horticultural Society, 1999 112 p40-43 2000

80. LINN (MC). "Cognition and Distance learning", Journal of the American Society for Information Science 47 p826-842 1996

81. LOWRY (CB). "Putting the pieces together: Essential Technologies for the virtual library" Journal of Academic Librarionship, p297 -300 Jul 1995

82. MCKIERMAN (G). "Give them what they want: participatory development of a WWW agricultural database collection", Quarterly Bulletin ofIAALD 42(3-4) p270-277 1997

83. MINCIONE (A); PUZONE (L) and HAZEKAMP (TH). "Internet and internet technologies applied to germplasm databases", Characterization and documentation of genetic resources utilizing multimedia databases proceedings p 17-21 1998

84. MISRA (BL). Universities without walls: The case of India, *Univ. News.* 26 (3) 1988; p. 3 - 5.

85. MITHCELL (S) and BALDWIN (CM). "INFOMINE; a unique virtual library of scholarly and educational resources in the agricultural and related sciences", Quarterly Bulletin ofIAALD 42(3-4) p230-233. 1997

86. MOHAN RAJAN (PA). The relevance of second and third laws of library science on the role of public libraries in commnunity information and distance education. *ILA Bulletin.23* (1-2) 1987. PP. 31 - 37.

87. MOHANTY (J). Educational broadcasting: Radio and television in education. Sterling publications, New Delhi.

88. MOORE (M). Learner autonomy: The second dimension of independent learning. *Convergence.*5, 1; 1972; p76-88.

89. NARAYANA REDDY(C). Andhra Pradesh Open University; Some reflections. *University News.* 24, 42; 1986; p17 – 20.

90. NIMONIY (S); KIURA (T); ETO (K); NANSEKI (T) and UEDA (M). "Field image acquisition system on WWW", Journal of Society of High Technology in Agriculture 9(1) 1997; p12-19

91. OJALA (Marydee). "What will they call us in the future?", Special liabraries 84(4) Fall 1993; p 226

92. PARMAJI (S). Distance education. New Delhi, Sterling, 1984. 29. Percival, Fred, *ed.,* Aspects of educational technology, Vol15 Distance learning and evaluation. London, Kogan page, 1981.

93. PERRY (W). Open University. Open University press, London. 1976.

94. PILLAI (AS). Instructional strategy for distance education. University News. 24(5), Feb. 1988. p.1 - 7.

95. PILLAI (Sivadasan K). Non-formal education in Britain: As an Indian adult education found it. Trivandrum, Kalaniketan, 1984.

96. POWELL (A)."Management Models and measurement in the virtual library", Special Libraries 85 (4) Fall 1994; p260-263

97. RAFFEL (JA) and SHISKO (Robert). Systematic analysis of university Libraries. MIT Press, Cambridge. 1969.

98. RAMAREDDY (G). Readings in distance education. IGNOU, New Delhi. 1987.

99. RANGANATHAN (SR). Academic library system: Fourth plan period. *LSD.* 2; 1965; p.293-347.

100. RASTOGI (Satish), *Ed.* Educational technology for distance education. Rawat publication, New Delhi. 1998.

101. ROWNTREE (D). Education technology in curriculum development. Harper and Raw, London.

102. SAHI (SS) and GUPTA (VK). Open University library. A conceptual analysis.ILA Bulletin. 21 (3-4) 1986; p.96-99.

103. SAHI (SS). Distance education system and library *science. Univ. News* 25 (11), March 1987; p. 6 - 8.

104. SARDANA (J L), *Ed.* Continuing education for librarians. ILA, New Delhi. 1984.

105. SCHMID (H); ZIEGLER (W) and POWELL (AP). "AGRIS and the internet", Quarterly bulletin of IAALD 41(1) 1996; p130-136

106. SEAGER (Daniel A). To maintain quality. Journal of education for librarianship. 4, 1; summer 1963; p.27 – 39.

107. SEATON (WJ). "Computer mediated communication and student self directed learning", Open learning 8(2) June 1993; p49-54

108. SEWART (David) *et. al. ed..* Distance education: International Perspectives. London, Croom Helm, 1983

109. SHARMA(S R). Organization of distance education. Pointer publisher, Jaipur. 1996.

110. SHERA (Jesse H.). The foundations of education for librarianship. Wiley Becker and Hayes publication, New York. 1972.

111. SINGH (Bakshish). Distance education: An experience of Open University. University News. 24, 6; 1986; p.21 – 24.

112. SINGH (Bakshshish). Distance education: An experience of Open University News. 24 (6), 1986; p.21 - 24.

113. SPRAGUE (J); DOERRY (E); DOUGLAS (S) and WESTERFIELD (M). "The Zebrafish Infonnation Network (ZFIN): a resource for genetic, genomic and developmental research", Nucleic Acids Research 29(1) 2001; p.87-90

114. Study material of Dr. B.R. Ambedkar Open University. BRAOU. Hyderabad.

115. Study material of Indira Gandhi National Open University. IGNOU. New Delhi.

116. Study material of Kota Open University. Kota Open University. Kota.

117. TAYLOR (JC) and WHITE (VJ). Why distnace education. Indian Journal of Adult Education. 47 (8), 1986.

118. TAYLOR (Robert S). Making of a library: The academic library in transition Wiley, New York. 1972.

119. TSUKIBOSHI (T) and SHIMANUKI (T). "WWW image database system: Encyclopedias of forage crop diseases", Grassland Science 45(2) 1999; p.193-197

120. Universities handbook of India. Association of Indian Universities, New Delhi. 1999.

121. University Grants Commission, “Guidelines for introduction of correspondence courses", Mimigraphed, 1981

122. VASHISHT (CP) *ed.,*. New Education policy Vis-a-Vis library development. Delhi, ILA, 1986.

123. VERMA (SC). New education policy and libraries. ILA Bulletin. 21 (3-4), 1985–86; p.100-106.

124. WACHTER (RM) and GUPTA (JND). "Distance education and the use of computers as instructional tools for system development projects: A case study of the Construction of Expert Systems", Computers & Education, 29 Jan 1997; p13-23

125. WALLDE (BV) and SCHILLER (N). "Creating the virtual Library: Strategic issues in the virtual library: Vision and realities" , West Port: CT:Mekler, 1993; p15-46

126. WHEELER (DL) et.al.. "Databases resources of the National Center for Biotechnology Information" , Nucleic Acids Research 29(1) 2001; p.11-16

127. ZHANG (P). "A Case study on technology uses in distance learning", Journal of Research on Computing in Education 30(4) Sum 1998

128. ZIGERELL (J). Distance education: An information age approach to adult education. Ohio, National Centre for Research in Vocational Education, 1984.

REFERENCES (WEB)

1. http://www.w3.org/2002/08/02-DOM-Disclosures.html
2. http://www.w3.org/2003/06/09-dom-core-issues/issues.html
3. http://www.w3.org/2002/10/DOM-Level-3-Val-issues
4. http://www.w3.org/2002/07/DOM-Level-3-Events-issues
5. http://www.w3.org/TR/2001/WD-DOM-Requirements-20010419/
6. http://lists.w3.org/Archives/Public/www-html-editor/
7. http://www.w3.org/MarkUp/Test/
8. http://lists.w3.org/Archives/Public/www-html-editor/
9. http://cgi.w3.org/MemberAccess/
10. http://www.rfc-editor.org/rfc/rfc2854.txt
11. http://lists.w3.org/Archives/Public/www-html/
12. http://www.w3.org/Talks/1999/05/www8-html/slide1.html
13. http://www.w3.org/Talks/1999/12/XHTML-XML99/slide1.html
14. http://www.gca.org/attend/1999_conferences/xml_99/
15. http://www.w3.org/2001/09/21-orf/xhtml-family/
16. http://www.kri.sfc.keio.ac.jp/ORF/2001/
17. http://www.w3.org/Talks/2002/04/11-pemberton
18. http://www.w3.org/Consortium/Offices/Germany/Events/Cross-Media-Publishing
19. http://www.w3.org/2002/Talks/www2002-xhtml/
20. http://www2002.org/> W3C Track http://www2002.org/w3ctrack.html
21. http://www.w3.org/2002/Talks/www2002-xhtml/Overview.xhtml
22. http://www.w3.org/2002/Talks/www2002-xhtml/Overview.html
23. http://www2003.org/> W3C Track http://www.w3.org/2003/03/w3c-track03.html
24. http://www.w3.org/2003/Talks/www2003-steven-horizontal/
25. http://www.w3.org/TR/WCAG10
26. http://www.w3.org/MarkUp/Guide/Style
27. http://www.boutell.com/faq/

28. http://lists.w3.org/Archives/Public/www-html/
29. http://lists.w3.org/Archives/Public/www-html-editor/
30. http://www.w3.org/2002/09/aa/
31. http://cgi.w3.org/MemberAccess/AccessRequest
32. http://lists.w3.org/Archives/Public/w3c-translators/
33. http://www.rfc-editor.org/rfc/rfc2557.txt
34. http://www.w3.org/MarkUp/HTML-WG
35. http://www2003.org/
36. http://tidy.sourceforge.net/
37. http://lists.w3.org/Archives/Public/html-tidy/
38. http://www.webreview.com/2000/06_16/webauthors/06_16_00_3.shtml
39. http://cgi.w3.org/cgi-bin/tidy
40. http://check.theinfo.org/html/tidy
41. http://valet.htmlhelp.com/tidy/
42. http://www.w3.org/People/Raggett/dtdgen/Docs/
43. http://www.w3.org/People/Raggett/tidy
44. http://www.w3.org/Talks/1999/03/24-stockholm-xhtml/
45. http://lists.w3.org/Archives/Public/www-math
46. http://camel.math.ca/mail/webmath/
47. http://www.dessci.com/en/support/tutorials/mathml/default.htm
48. http://www.charlesriver.com/titles/mathml.html
49. http://www.w3.org/People/maxf/papers/iamc.ps
50. http://www.mathtype.com/en/reference/webmath/status/status_Sep_02.htm
51. http://www.mathweb.org
52. http://www.ams.org/mathweb
53. http://www.mathmlcentral.com/
54. http://www.dessci.com/en/reference/webmath/tech/mathml.htm
55. http://www.oasis-open.org/cover/topics.html
56. http://www.zvon.org/HowTo/Output/index.html
57. http://pear.math.pitt.edu/mathzilla/

58. http://www.activemath.org/
59. http://mathforum.org/
60. http://www.joma.org/
61. http://www.w3.org/TR/2003/WD-MathML2-20030411/
62. http://www.w3.org/Consortium/Process-20010719/tr.html
63. http://www.w3c.or.kr/Translation/Math/XSL/Overview-Korean.html
64. http://www.soft4science.com
65. http://www.charlesriver.com/titles/mathml.html
66. http://www-dft.ts.infn.it/mathml/
67. http://mathosphere.net/editeurml/doc_english.html
68. http://mathosphere.net/editeurml/WeM.html
69. http://mathosphere.net/outils/index.html
70. http://www.w3.org/Math/testsuite
71. http://www.dessci.com/webmath/mathplayer
72. http://www.w3.org/Math/XSL/Overview-Chinese.html
73. http://eenk.new21.org/Math/XSL/Overview-Korean.html
74. http://www.w3.org/Math/XSL/pmathml2.xml
75. http://www.w3.org/Math/XSL/csmall2.xml
76. http://www.w3.org/TR/MathML2/chapter3.html
77. http://www.w3.org/Math/charter.HTML
78. http://www.w3.org/TR/REC-MathML/chapter2.HTML
79. http://www.w3.org/Math/Group/needs/MathMLRequ.html
80. http://www.w3.org/MarkUp/CoordGroup
81. http://www.w3.org/Math/Group/needs/MathMLRequ.html
82. http://www.adobe.com/
83. http://www.ams.org/
84. http://www.mathtype.com/mathtype/
85. http://www.elsevier.nl/
86. http://www.webeq.com/
87. http://www.ics.raleigh.ibm.com/ics/techexp.htm

88. http://www.mathsoft.com
89. http://www.uwaterloo.ca/
90. http://www.inria.fr/Equipes/SAFIR-eng.HTML
91. http://www.stilo.com
92. http://www.sq.com/
93. http://www.csd.uwo.ca/
94. http://www.maplesoft.com/
95. http://www.wri.com/
96. http://www.mathtype.com
97. http://www.w3.org/MarkUp/CoordGroup/9802/xml-in-html
98. http://www.w3.org/MarkUp/future/
99. http://www.w3.org/TR/WD-math/
100. http://www.dessci.com/webmath/mathplayer
101. http://www.nag.co.uk/projects/OpenMath/mml-files/
102. http://www.w3.org/People/Raggett/EzMath/
103. http://www.mathcad.com
104. http://www.wolfram.com
105. http://www.newmexico.mackichan.com/mathml/mathmled.htm
106. http://www.mathmlconference.org/2002/tutorials.html
107. http://www.w3.org/Math
108. http://www.mathmlconference.org/2002
109. http://www.adobe.com/acrofamily/main.html
110. http://www.dessci.com/en/reference/webmath/status/ </en/reference/webmath/status/
111. http://www.w3.org
112. http://www.dessci.com/en/products/mathplayer/ </en/products/mathplayer/
113. http://www.microsoft.com/windows/ie/default.asp
114. http://www.dessci.com/en/products/webeq/ </en/products/webeq/
115. http://www.mozilla.org
116. http://channels.netscape.com/ns/browsers/7/default.jsp

117. http://www.w3.org/Style/XSL/
118. http://www.w3.org/Math/XSL/
119. http://www.dessci.com/en/reference/webmath/strategies.htm
120. http://www.mathmlconference.org/2002/tutorials.html
121. http://www.w3.org/TR/MathML2/chapter2.html#fund_overview
122. http://cnx.rice.edu/
123. http://www.ams.org/msc/
124. http://www.openmath.org/
125. http://www.wolfram.com
126. http://www.maplesoft.com
127. http://www.mowgli.cs.unibo.it
128. http://www-2.cs.cmu.edu/~ccaps/
129. http://www.dessci.com/en/products/mathtype/ </en/products/mathtype/
130. http://www.mathmlcentral.com/
131. http://www.ecollege.com/stories/press_08_14_02.html
132. http://www.ecollege.com/stories/press_05_15_02.html
133. http://jigsaw.w3.org/new.html
134. http://lists.w3.org/Archives/Public/www-jigsaw/
135. http://jigsaw.w3.org/
136. http://www.latoserver.it/jigsaw/
137. http://www.w3.org/pub/WWW/COPYRIGHT.html
138. http://dev.w3.org/cvsweb/~checkout~/java/classes/org/w3c/jigsaw/https/readme.txt?rev=1.1&content-type=text/plain
139. http://jigsaw.w3.org/Winie/
140. http://www.w3.org/1999/04/Editing/
141. http://www.webdav.org/
142. http://lists.w3.org/Archives/Public/www-jigsaw/
143. http://jigsaw.w3.org/Doc/JigsawDoc.pdf
144. http://jigsaw.w3.org/Doc/JigsawDoc.ps.gz
145. http://jigsaw.w3.org/Distrib/jigsaw_2.2.2.zip

146. http://jigsaw.w3.org/Distrib/jigsaw_2.2.2.tar.gz
147. http://jigsaw.w3.org/Distrib/jigsaw_2.2.2.tar.bz2
148. http://jigsaw.w3.org/Devel/Mirror/jigsaw_2.0.5.zip
149. http://jigsaw.w3.org/Devel/Mirror/jigsaw_2.0.5.tar.gz
150. http://jigsaw.w3.org/Devel/Mirror/jigsaw_2.0.5.tar.bz2
151. http://jigsaw.w3.org/Devel/Mirror/jigsaw_webdav_2.1.2.zip
152. http://jigsaw.w3.org/Devel/Mirror/jigsaw_webdav_2.1.2.tar.gz
153. http://jigsaw.w3.org/Devel/Mirror/jigsaw_webdav_2.1.2.tar.bz2
154. ftp://ftp.w3.org/pub/jigsaw/jigsaw.zip
155. ftp://ftp.w3.org/pub/jigsaw/jigsaw.tar.gz
156. http://www.w3.org/TR/2002/NOTE-photo-rdf-20020419
157. http://www.w3.org/TR/photo-rdf
158. http://www.w3.org/TR/2000/NOTE-photo-rdf-20000928/
159. http://www.w3.org/Consortium/Legal/ipr-notice-20000612
160. http://www.w3.org/Consortium/Legal/ipr-notice-20000612
161. http://www.w3.org/Consortium/Legal/ipr-notice-20000612
162. http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary/
163. http://lists.w3.org/Archives/Public/www-smil/
164. http://www.roxia.co.kr
165. http://www.alphaworks.ibm.com/tech/tk4mpeg4
166. http://www.psych.uiuc.edu/~kmiller/dvguide/analysis_tools.htm
167. http://www.psych.uiuc.edu/~kmiller/smil/smil_umbrella.htm
168. http://www.smilmedia.com/
169. http://k-tai.impress.co.jp/cda/article/news_toppage/13103.html
170. http://developer.openwave.com/technotes/client_mms_capabilities.html
171. http://www.w3.org/TR/smil20
172. http://www.smilmedia.com/spec/spec1~7.htm
173. http://www.w3.org/TR/XHTMLplusSMIL/
174. http://www.w3.org/2001/SMIL20/testsuite/
175. http://www.w3.org/2001/05/23/SMIL-Implementation-result.html

176. http://www.mobilemms.com/mmsfaq.asp
177. http://lightning.prohosting.com/~qqiu/smil/trans/REC-smil-19980615-cn.html
178. http://www.sunshine-company.de/w3c/REC-smil-19980615-DE.html
179. http://www.w3c.cnr.it/office/traduzioni/REC-smil-it.html
180. http://www.doraneko.org/misc/smil10/smil10.html
181. http://www.mentallink.com/resource/smil/smil10-kr.html
182. http://www.utad.pt/~leonelm/w3ctranslations/smil/
183. http://smil.nist.gov/Testcase.html
184. http://smil.nist.gov/Feature.html
185. http://service.real.com/help/library/guides/production/realpgd.htm
186. http://developer.apple.com/techpubs/quicktime/qtdevdocs/PDF/insideqt_intmov.pdf
187. http://www.computer.org/multimedia/mu2001/pdf/u4082.pdf
188. http://www.xml.com/pub/a/2002/05/29/smil.html
189. http://www.allhtml.com/smil/index.php
190. http://www.smilbook.com/
191. http://www.w3.org/Talks/2001/06CSMIL/slide1-0.html
192. http://developer.openwave.com/technotes/client_mms_capabilities.html
193. http://www.webtechniques.com/archives/1998/09/bouthillier/
194. http://www.cwi.nl/~media/SMIL/Tutorial/
195. http://www.helio.org/products/smil/tutorial/
196. http://www.empirenet.com/~joseram/index.html
197. http://webreview.com/wr/pub/1999/03/12/feature/index.html
198. http://www.kevlindev.com/basics/index.htm
199. http://www.kevlindev.com/
200. http://www.empirenet.com/~joseram/universal/universal.html
201. http://aristote1.aristote.asso.fr/Presentations/Smil/Courtaud/SMIL-Presentation-fr.smi
202. http://www.euroclid.fr/Cours_SMIL_W3C/
203. http://v.hbi-stuttgart.de/~keitz/skripte/SMILStart.htm
204. http://lists.w3.org/Archives/Public/www-smil/

205. http://realforum.real.com/cgi-bin/realforum/postlist.pl?Cat=&Board=ccsmil
206. http://www.real.com
207. http://www.html.it/smil/index.html
208. http://www.realnetworks.com/solutions/ecosystem/realone.html?src=rnhmfs
209. http://www.oratrix.com/GRiNS/SMIL-2.0/
210. http://www.inobject.com/mmplay.htm
211. http://www.inobject.com/InterObjectSMILPlayerSpecification.htm
212. http://www.microsoft.com/windows/ie/preview/default.asp
213. http://www.w3.org/TR/2001/WD-XHTMLplusSMIL-20010807/
214. http://www.microsoft.com/windows/ie/default.htm
215. http://msdn.microsoft.com/workshop/Author/behaviors/htmltime.asp
216. http://k-tai.impress.co.jp/cda/article/news_toppage/13103.html
217. http://wam.inrialpes.fr/software/pocketsmil/
218. http://www.roxia.co.kr
219. http://lists.w3.org/Archives/Public/www-mobile/2002Aug/0007.html
220. http://www.oratrix.com/GRiNS/index.html
221. http://www.research.digital.com/SRC/HPAS
222. http://www.prodworks.com/
223. http://www.apple.com/quicktime/authoring/qtsmil.html
224. http://www.real.com/
225. http://www.helio.org
226. http://smil.nist.gov/player
227. http://www.salzburgresearch.at/suntrec/schmunzel/
228. http://www.xsmiles.org/
229. http://www.smilmedia.com
230. http://www.confluenttechnologies.com
231. http://www.oratrix.com/GRiNS/index.html
232. http://www.adobe.com/products/golive/overview.html
233. http://www.allaire.com/products/homesite/index.cfm
234. http://ncam.wgbh.org/webaccess/magpie

235. http://www.libpng.org/pub/png/
236. http://www.libpng.org/pub/png/pngdocs.html
237. http://www.w3.org/Graphics/PNG/
238. http://www.libpng.org/pub/png/pngapps.html
239. http://www.jpeg.org/
240. http://www.iso.ch/
241. http://www.jpeg.org/CDs15444.html
242. http://www.lizardtech.com/solutions/geospatial/
243. http://www.lizardtech.com/solutions/document/
244. http://www.djvu.com/solutions/document/whitepaper/
245. http://www.djvuzone.org/
246. http://www.planetdjvu.com/
247. http://djvu.sourceforge.net/
248. http://www.lizardtech.com/
249. http://www.celartem.com/
250. http://www.extensis.com/pxlsmartscale/
251. http://www.hisoftware.com/hmcc/acc4mcc.html
252. http://www.simple.co.jp/products/10MovieBorad.htm
253. http://lists.w3.org/Archives/Public/www-mobile/2002Aug/0007.html
254. http://www.webiphany.com/perlysmil/
255. http://forms.real.com/rnforms/products/tools/slideshowbasic/index.html?key=868E21032182964
256. http://autodownload.sausage.com
257. http://smibase.com/
258. http://www.docomo-sys.co.jp/prod/soft/smil2.html
259. http://www.smilgen.org
260. http://www.smilme.com
261. http://w3-mcgav.lab.kdd.co.jp/sc/indexe.html
262. http://tag.digital-ren.com
263. http://www.tagfree.com/english/product/product02.asp?menu=2

264. http://www.psych.uiuc.edu/~kmiller/dvguide/analysis_tools.htm
265. http://www.veon.com/
266. http://www.cwi.nl/~media/symm/validator/
267. http://www.smilmedia.com.
268. http://www.alphaworks.ibm.com/tech/tk4mpeg4
269. http://www.justsmil.com
270. http://www.wgbh.org/wgbh/pages/ncam/webaccess/captionedmovies.html
271. http://www.w3.org/TR/SVG/
272. http://www.w3.org/TR/1999/WD-SVG-19990211
273. http://www.oreilly.com/catalog/learnxml/index.html?CMP=IL7015
274. http://www.oreilly.com/catalog/learnxml/index.html?CMP=IL7015
275. http://www.oreillynet.com/cs/catalog/view/au/752?x-t=book.view&CMP=IL7015
276. http://www.oreilly.com/catalog/learnxml/toc.html?CMP=IL7015
277. http://www.oreilly.com/catalog/learnxml/inx.html?CMP=IL7015
278. http://www.oreilly.com/catalog/learnxml/chapter/index.html?CMP=IL7015
279. http://xml.oreilly.com/news/learningxml_0101.html?CMP=IL7015
280. http://safari.oreilly.com/0596000464
281. http://www.adobe.com/svg/viewer/install/
282. http://www.adobe.com/svg/
283. http://www.croczilla.com/SVG
284. http://www.xmlspy.com/
285. http://www.w3.org/TR/SVG11/
286. http://www.w3.org/TR/SVG/
287. http://www.w3.org/TR/SVGMobile/
288. http://www.w3.org/TR/SVG12
289. http://www.w3.org/Graphics/SVG/
290. http://www.w3.org/Graphics/SVG/Test/
291. http://www.peoplesnetwork.gov.uk/nof/technicalstandards.html/
292. http://heds.herts.ac.uk/resources/costing.html/
293. http://heds.herts.ac.uk/resources/matrix.html/

294. http://www.rlg.ac.uk/preserv/diginews/diginews3-5.html/
295. http://www.leeds.ac.uk/cedars/
296. http://www.bl.uk/services/preservation/digcult.html/
297. http://heds.herts.ac.uk/resources/papers/jidi_fs.html/
298. http://www.tasi.ac.uk/framework/capture/image.html/
299. http://www.epi-centre.com/basics/basics2.html/
300. http://www.image-acquire.com/scanner/
301. http://www.rlg.org/preserv/diginews/diginews3.html#hardware&software
302. http://www.rlg.org/preserv/diginews/diginews3.html
303. http://www.imaging-resource.com/
304. http://www.epi-centre.com/reports/reports.html/
305. http://www.tasi.ac.uk/framework/capture/camera.html/
306. http://www.cimtech.co.uk/cimtech/pub/p1.htm/
307. http://www.jstor.org/
308. http://imago.library.mcgill.ca/osler
309. http://www.xerox.com/scansoft/pagis
310. http://antonio.emeraldinsight.com/rpsv/cgi-bin/linker?ext=y&ref=23817bb1-3
311. http://computer.org/ computer/dli/index.html/
312. http://www.nlc-bnc.ca/cidl/cidle.html/
313. http://www.dlib.org/dlib/july97/07chesnutt.tml
314. http://antonio.emeraldinsight.com/rpsv/cgi-bin/linker?ext=y&ref=23817bb1-7
315. http://www.gii. getty.du/index/imaginit.html
316. http://www.mellon.org/jsback.tm
317. http://www.jstor.org/about/production.html
318. http://www.princeton.edu/cit/campus_com/jstorsep
319. http://www.nlc-bnc.ca/pubs/netnotes/notes14
320. http://www.princeton.edu/cit/campus_com/jstor2.tml
321. http://www.rlg.org/preserv/
322. http://www.geom.umn.du/docs/cecm/steinberger/talk.html
323. http://www.orbeon.com/oxf/doc/processors-xforms

324. http://www.alphaworks.ibm.com/tech/xmlforms
325. http://www.xmlform.org
326. http://www.formsplayer.com/
327. http://www.xsmiles.org/features_xforms.html
328. http://www.novell.com/xforms
329. http://www.w3.org/TR/2002/CR-xforms-20021112/
330. http://www.w3.org/2002/11/xforms-pressrelease
331. http://www.w3.org/2002/11/xforms-cr-testimonial.html
332. http://www.w3.org/TR/2002/CR-xforms-20021112/http://lists.w3.org/Archives/Public/www-forms-editor/
333. http://www.FormsPlayer.com/
334. http://www.novell.com/xforms
335. http://trustform.comsquare.co.kr
336. http://www.achieveforms.com
337. http://jxforms.cybernd.at/
338. http://www.handwise.com/Products/XFUI.html
339. http://www.holosofx.com/x_form.htm
340. http://sourceforge.net/projects/xserverforms
341. http://www.alphaworks.ibm.com/tech/xmlforms
342. http://www.orbeon.com/oxf/doc/processors-xforms
343. http://www.formfaces.com/
344. http://www.nforms.net/
345. http://www.w3.org/TR/xforms
346. http://www.cardiff.com/XForms
347. http://www.w3.org/TR/1999/REC-xpath-19991116
348. http://www.w3.org/TR/1999/REC-xpath-19991116.xml
349. http://www.w3.org/TR/1999/REC-xpath-19991116.html
350. http://www.w3.org/TR/xpath
351. http://www.w3.org/TR/1999/PR-xpath-19991008
352. http://www.w3.org/1999/08/WD-xpath-19990813

353. http://www.w3.org/1999/07/WD-xpath-19990709
354. http://www.w3.org/TR/1999/WD-xslt-19990421
355. http://www.w3.org/1999/11/REC-xpath-19991116-errata http://lists.w3.org/Archives/Public/www-xpath-comments
356. http://validator.w3.org
357. http://validator.w3.org/about.html
358. http://validator.w3.org/source/
359. http://validator.w3.org/checklink
360. http://www.w3.org/2000/07/checklink
361. http://dev.w3.org/cvsweb/validator/httpd/cgi-bin/checklink.pl
362. http://jigsaw.w3.org/css-validator/
363. http://jigsaw.w3.org/css-validator/README
364. http://jigsaw.w3.org/css-validator/DOWNLOAD.html
365. http://www.w3.org/RDF/Validator/>
366. http://www.w3.org/P3P/validator
367. http://dev.w3.org/cvsweb/p3p-validator/20020128/
368. http://www.w3.org/QA/Tools/LogValidator/
369. http://www.w3.org/QA/Tools/MUTAT/
370. http://www.w3.org/Protocols/HTTP/Forum/
371. http://lists.w3.org/Archives/Public/www-html/1994Jul/0015
372. http://www.w3.org/QA/Tools/LogValidator/Manual
373. http://www.w3.org/P3P/compliant_sites
374. http://www.w3.org/P3P/validator/20000510/
375. http://www.w3.org/P3P/validator/20020128/p3p.xml
376. http://www.w3.org/P3P/validator/20020128/base.xml
377. http://dev.w3.org/cvsweb/p3p-validator/20020128/
378. http://www.w3.org/TR/REC-CSS2/
379. http://lists.w3.org/Archives/Member/w3c-voice-wg/2003Feb/0084.html
380. http://cgi.w3.org/MemberAccess/AccessRequest
381. http://lists.w3.org/Archives/Public/www-voice/

382. http://www.voicexml.org/
383. http://www.speechtechmag.com/
384. http://www.kenrehor.com/voicexml/
385. http://www.speech.cs.cmu.edu/openvxi/
386. http://www.heyanita.com/
387. http://www-3.ibm.com/software/pervasive/products/voice/voice_technologies.shtml
388. http://www.intervoice.com/
389. http://www.loquendo.com/
390. http://www.loquendocafe.com/
391. http://fife.speech.cs.cmu.edu/openvxi/
392. http://www.speechworks.com/
393. http://www.pipebeach.com
394. http://www.publicvoicexml.org/
395. http://sourceforge.net/projects/publicvoicexml/
396. http://www.speechworks.com/
397. http://www.voicegenie.com/
398. http://developer.voicegenie.com
399. http://www.voxpilot.com/
400. http://ode.voxpilot.com/
401. http://www.microsoft.com/speech/
402. http://www.w3.org/WAI/GL/2003/07/f2f-agenda.html
403. http://www.w3.org/WAI/UA/impl-pr2/
404. http://www.w3.org/WAI/UA/TS/
405. http://www.w3.org/WAI/UA/2002/08/eval
406. http://www.w3.org/WAI/events.html
407. http://www.w3.org/WAI/events.html
408. http://www.w3.org/WAI/about.html
409. http://www.w3.org/WAI/about.html
410. http://www.w3.org/WAI/about.html
411. http://www.w3.org/WAI/contacts.html

412. http://www.cgmopen.org/

413. http://www.cgmopen.org/webcgmintro/paper.htm

414. http://www.cgmopen.org/

415. http://www.w3.org/2001/12/xmlbp/xml-linking-wg-charter.html

416. http://www.w3.org/TR/xlink/

417. http://www.w3.org/TR/xmlbase/

418. http://www.w3.org/2001/06/link-base-pressrelease

419. http://www.sun.com/software/xml/developers/xlink.html

420. http://www.xmlsoftware.com/xlink/

421. http://www.simonstl.com/projects/xlinkfilter/resources.html

422. http://www.zvon.org/xxl/xlink/Output/xlink_refs.html

423. http://www.zvon.org/xxl/xlink/xlink_extend/OutputExamples/frame_xlinkextend_html.html

424. http://www.w3.org/TR/2003/REC-xptr-framework-20030325/

425. http://www.cs.unibo.it/~fabio/XPointer/

426. http://www.google.com/search?q=XML+Conferences

427. http://www.google.com/search?q=XML+books

428. http://www.google.com/search?q=XML+training+courses

429. http://www.google.com/search?q=XML+online+tutorials

430. http://www.google.com/search?q=XML+bibliography

431. http://www.google.com/search?q=XML+parser

432. http://www.google.com/search?q=XML+magazines

433. http://www.google.com/search?q=%22XML+movie%22

434. http://www.w3.org/TR/2001/REC-xmlbase-20010627/

435. http://www.w3.org/TR/2001/REC-xmlbase-20010627/Overview.html

436. http://www.w3.org/TR/2001/REC-xmlbase-20010627/Overview.xml

437. http://www.w3.org/TR/xmlbase/

438. http://www.w3.org/TR/2000/PR-xmlbase-20001220/

439. http://www.w3.org/XML/Activity

440. http://www.w3.org/2001/06/xmlbase-translations

441. http://lists.w3.org/Archives/Public/xml-encryption/
442. http://www.w3.org/TR/xml-encryption-req
443. http://www.w3.org/TR/xmlenc-decrypt
444. http://lists.w3.org/Archives/Public/xml-encryption/
445. http://www.w3.org/2000/09/XML-Encryption-Workshop.html
446. http://www.w3.org/Signature/2001/04/05-xmldsig-interop.html
447. http://www.w3.org/TR/xmldsig-core/
448. http://jcewww.iaik.tu-graz.ac.at/products/ixsil/index.php
449. http://www.ubisecure.com/index.php?page=ubisignature
450. http://www.xmltrustcenter.org/xmlsig/developer/verisign/index.htm
451. http://www.wedgetail.com/xmlsecurity/
452. http://www.w3.org/TR/xmldsig-requirements
453. http://www.w3.org/TR/1999/REC-xpath-19991116.xml
454. http://www.w3.org/TR/1999/REC-xpath-19991116.html
455. http://www.w3.org/TR/xpath
456. http://www.w3.org/TR/1999/PR-xpath-19991008
457. http://www.w3.org/1999/08/WD-xpath-19990813
458. http://www.w3.org/1999/07/WD-xpath-19990709
459. http://www.w3.org/TR/1999/WD-xslt-19990421

QUESTIONNAIRE

Directions: The question below ask for information about yourself and for your opinions regarding various aspects of part time/distance study in different programs being offered by Indian Open Universities. Please answer all the questions given below. On all multiple-choice items, tick only appropriate responses.

1. To which Open University do you belong?

--Y. Chavhan Mharashtra Open University --Kota Open University

--Indira Gandhi National Open University --Any other

(Pls. Specify--------------------------)

2. Level of programe in which you are enrolled?

--Certificate Course --Diploma Course

--Bachelor Course --Master Course

3. Your sex is

--Male --Female

4. Your age is between

--18 years or younger --19 years to 28 years

--29 years to 35 years --36 years or more

5. Where do you live (Permanent resident)?

State-------------------- District----------------------

City--------------------- Village----------------------

6. Are you studying while working at the same time?

--Yes --No

If yes, please specify your designation--

7. How many years work experience has you had?

--None
--Up to 2 years
--2 to 5 years
--More than 5 years

8 Why are you doing this course?

--For adding qualification
--For getting a professional degree
--For departmental promotion
--Any other reason
(Pls. Specify-------------------------)

9. What is the main reason you decided to enroll as a part time/distant student rather than a regular student?

--In-service study
--Could not get admission
--Family comitments
--Any other
(Pls. Specify-------------------------)

10. What is the main reason you decided to apply to this open university's part time programe?

--Reputation of the programe
--Low fee structure
--Only programe in geographical area
--Any other (Pls. Specify-------------------------
-------------------------)

11. What is the medium of instruction during interaction?

--English
--Local language
--Hindi
--Any other language
(Pls. Specify-------------------------)

12. Through which type of educational technology do you receive instructions?

--Print media
--Audio/Video cassettes
--Radio
--Television
--Telephone
--E-Mail

13. Throughout your program, how easy it has been to interact/communicate with your instructors/teachers?

--Very easy --Easy

--Difficult --Very difficult

14. What is the main medium you use to communicate with your instructors/teachers?

--Face-to-face --Telephone

--E-Mail --Any other

(Pls. Specify---------------------------)

15. On average per week, how many other students of your class do you communicate outside the class?

--None --1 to 3

--4 to 7 --8 or more

16. What is the main medium you use to communicate with other students out side the class?

--Face-to-face --Telephone

--E-Mail --Any other

(Pls. Specify--------------------------)

17. Do you physically access library services of your open university?

--Yes --No

If no, then what is the reason behind it?

--Long distance --Not allowed

--Don't have interest --Not needed at all

8 While studying in the specified course have you ever tried to access the Web Site of your open university

--Yes --No

If yes, do you feel that university should provide library access through web

--Yes --No

If yes, which type of web based library service do you need?

--Indexing --Abstracting

--OPAC --CAS

--SDI --E-Consultation

--Full-text search --Any other

(Pls. Specify------------------------)

If web based full text service then which type of document do you want to access full text?

--Reference document --Text document

--Periodical --Any other

(Pls. Specify------------------------)

19. Do you have any suggestion regarding any web-based service of your library or open university?

--Yes --No ✓

If yes, please specify--

----

20 Do you face any specific problem during this course work?

--Yes --No

If yes, please specify--
